

---

# **BACHELORARBEIT**

---

Herr  
**Stefan Pabst**

**Entwicklung eines Goldstandards  
zur Evaluation von Algorithmen  
für die Detektion von Gefährdern  
in sozialen Netzwerken**

2017



# **BACHELORARBEIT**

---

## **Entwicklung eines Goldstandards zur Evaluation von Algorithmen für die Detektion von Gefährdern in sozialen Netzwerken**

Autor:

**Stefan Pabst**

Studiengang:

Allgemeine und Digitale Forensik

Seminargruppe:

FO14w1-B

Erstprüfer:

Prof. Dr. rer. nat. Dirk Labudde

Zweitprüfer:

Michael Spranger, M. Sc.

Mittweida, September 2017



---

## **Bibliografische Angaben**

Pabst, Stefan: Entwicklung eines Goldstandards zur Evaluation von Algorithmen für die Detektion von Gefährdern in sozialen Netzwerken, 121 Seiten, 9 Abbildungen, Hochschule Mittweida, University of Applied Sciences, Fakultät Angewandte Computer- und Biowissenschaften

Bachelorarbeit, 2017

## **Referat**

Die Überwachung sozialer Online-Netzwerke spielt eine zunehmend wichtige Rolle, um die polizeilichen Aufgaben hinsichtlich Kriminalprävention und Gefahrenabwehr auch im virtuellen Raum effektiv und effizient wahrnehmen zu können. Gleichzeitig machen das enorme Volumen und die Komplexität der Daten eine manuelle Bearbeitung nahezu unmöglich. Es müssen automatisierte Verfahren entwickelt werden, welche sogenannte Gefährder in sozialen Online-Netzwerken sicher erkennen können. Für die Entwicklung, das Training und die Evaluation von Modellen für das Maschinelle Lernen werden standardisierte Referenzdaten benötigt. Die vorliegende Arbeit beschäftigt sich mit der Entwicklung eines solchen Goldstandards mit Trainings- und Testdaten für Algorithmen zur Detektion von Gefährdern in sozialen Online-Netzwerken. Es werden Kriterien für die Auswahl relevanter Profile und Attribute erarbeitet und Anforderungen für die Strukturierung und Ablage der Daten formuliert. Weiterhin werden konkrete Profile und Feature-Kandidaten sowie ein XML-Schema und Dateilayout für die Bereitstellung der Daten des Goldstandards vorgeschlagen. Abschließend werden die Ergebnisse kritisch gewürdigt und ein Ausblick für zukünftige Arbeiten gegeben.



# I. Inhaltsverzeichnis

<b>Inhaltsverzeichnis</b>	<b>I</b>
<b>Abbildungsverzeichnis</b>	<b>II</b>
<b>Tabellenverzeichnis</b>	<b>III</b>
<b>Abkürzungsverzeichnis</b>	<b>IV</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Zielstellung . . . . .	3
1.3 Aufbau der Arbeit . . . . .	3
<b>2 Analyse sozialer Netzwerke</b>	<b>5</b>
2.1 Exkurs: Graphentheorie . . . . .	5
2.2 Beziehungen in sozialen Netzwerken . . . . .	7
2.3 Soziale Online-Netzwerke . . . . .	8
2.3.1 Exkurs: Social Media . . . . .	8
2.3.2 Merkmale und Funktionen . . . . .	10
2.4 Forschungsobjekt Facebook . . . . .	11
2.4.1 Facebooks Social Graph . . . . .	13
2.4.2 Anatomie einer Facebook Page . . . . .	15
<b>3 Detektion von Gefährdern</b>	<b>21</b>
3.1 Merkmale von Gefährdern . . . . .	21
3.2 Exkurs: Klassifizierung und Maschinelles Lernen . . . . .	23
3.2.1 Finden geeigneter Features . . . . .	24
3.2.2 Wahl des Klassifizierungsverfahrens . . . . .	26
3.2.3 Kombinieren von Klassifikatoren . . . . .	27
3.2.4 Evaluation von Klassifizierungsalgorithmen . . . . .	28
3.3 Herausforderungen bei der Klassifizierung im Kontext von Social Media . . . . .	30
<b>4 Entwicklung des Goldstandards</b>	<b>33</b>
4.1 Anforderungen . . . . .	33
4.1.1 Relevante Profile . . . . .	34
4.1.2 Relevante Attribute . . . . .	34
4.1.3 Abgrenzung der Daten . . . . .	35
4.1.4 Bereitstellung der Daten . . . . .	38
4.2 Auswahl geeigneter Profile . . . . .	40
4.3 Auswahl von Feature-Kandidaten . . . . .	42
4.4 Struktur und Format der Daten . . . . .	44
<b>5 Fazit und Ausblick</b>	<b>49</b>

<b>A</b>	<b>Fields und Edges relevanter Facebook Nodes</b>	<b>51</b>
A.1	Album Node . . . . .	51
A.2	Comment Node . . . . .	53
A.3	Event Node . . . . .	55
A.4	Page Node . . . . .	57
A.5	Photo Node . . . . .	68
A.6	Picture Node . . . . .	70
A.7	Post Node . . . . .	71
A.8	User Node . . . . .	75
A.9	Video Node . . . . .	82
A.10	VideoList Node . . . . .	84
<b>B</b>	<b>Feature-Kandidaten relevanter Facebook Nodes</b>	<b>85</b>
B.1	Album Node . . . . .	85
B.2	Comment Node . . . . .	87
B.3	Event Node . . . . .	88
B.4	Page Node . . . . .	90
B.5	Photo Node . . . . .	93
B.6	Picture Node . . . . .	94
B.7	Post Node . . . . .	95
B.8	User Node . . . . .	97
B.9	Video Node . . . . .	98
B.10	VideoList Node . . . . .	99
<b>C</b>	<b>Facebook-Profile relevanter Gruppierungen in Sachsen</b>	<b>101</b>
<b>D</b>	<b>XML-Schema des Goldstandards</b>	<b>103</b>
	<b>Literaturverzeichnis</b>	<b>109</b>



---

## II. Abbildungsverzeichnis

2.1 Page Node mit Edges und potenziell verbundenen Nodes (1-Schritt-Umgebung) . . .	18
2.2 Page Node mit Edges und potenziell verbundenen Nodes (2-Schritt-Umgebung) . . .	19
3.1 Allgemeiner Prozess der Klassifizierung . . . . .	24
3.2 Feature Selection im Klassifizierungsprozess . . . . .	25
3.3 Feature Transformation für bessere lineare Trennbarkeit . . . . .	26
3.4 Schematische Struktur eines künstlichen Immunsystems . . . . .	28
4.1 Exemplarischer Screenshot mit den Elementen einer Facebook Page . . . . .	37
4.2 Struktur des XML-Schemas für die Abbildung der Profildaten des Goldstandards (1/2)	46
4.3 Struktur des XML-Schemas für die Abbildung der Profildaten des Goldstandards (2/2)	47



## III. Tabellenverzeichnis

2.1	Allgemein verfügbare Fields eines Page Node . . . . .	16
2.2	Allgemein verfügbare Edges eines Page Node . . . . .	17
4.1	Feature-Kandidaten der Fields eines Page Node . . . . .	43
4.2	Feature-Kandidaten der Edges eines Page Node . . . . .	44
A.1	Fields eines Album Node . . . . .	51
A.2	Edges eines Album Node . . . . .	52
A.3	Fields eines Comment Node . . . . .	53
A.4	Edges eines Comment Node . . . . .	54
A.5	Fields eines Event Node . . . . .	55
A.6	Edges eines Event Node . . . . .	56
A.7	Fields eines Page Node . . . . .	57
A.8	Edges eines Page Node . . . . .	64
A.9	Fields eines Photo Node . . . . .	68
A.10	Edges eines Photo Node . . . . .	69
A.11	Fields eines Picture Node . . . . .	70
A.12	Edges eines Picture Node . . . . .	70
A.13	Fields eines Post Node . . . . .	71
A.14	Edges eines Post Node . . . . .	74
A.15	Fields eines User Node . . . . .	75
A.16	Edges eines User Node . . . . .	78
A.17	Fields eines Video Node . . . . .	82
A.18	Edges eines Video Node . . . . .	84
A.19	Fields eines VideoList Node . . . . .	84
A.20	Edges eines VideoList Node . . . . .	84
B.1	Feature-Kandidaten der Fields eines Album Node . . . . .	85
B.2	Feature-Kandidaten der Edges eines Album Node . . . . .	86
B.3	Feature-Kandidaten der Fields eines Comment Node . . . . .	87
B.4	Feature-Kandidaten der Edges eines Comment Node . . . . .	87

B.5	Feature-Kandidaten der Fields eines Event Node . . . . .	88
B.6	Feature-Kandidaten der Edges eines Event Node . . . . .	89
B.7	Feature-Kandidaten der Fields eines Page Node . . . . .	90
B.8	Feature-Kandidaten der Edges eines Page Node . . . . .	92
B.9	Feature-Kandidaten der Fields eines Photo Node . . . . .	93
B.10	Feature-Kandidaten der Edges eines Photo Node . . . . .	94
B.11	Feature-Kandidaten der Fields eines Picture Node . . . . .	94
B.12	Feature-Kandidaten der Edges eines Picture Node . . . . .	94
B.13	Feature-Kandidaten der Fields eines Post Node . . . . .	95
B.14	Feature-Kandidaten der Edges eines Post Node . . . . .	97
B.15	Feature-Kandidaten der Fields eines User Node . . . . .	97
B.16	Feature-Kandidaten der Edges eines User Node . . . . .	97
B.17	Feature-Kandidaten der Fields eines Video Node . . . . .	98
B.18	Feature-Kandidaten der Edges eines Video Node . . . . .	98
B.19	Feature-Kandidaten der Fields eines VideoList Node . . . . .	99
B.20	Feature-Kandidaten der Edges eines VideoList Node . . . . .	99
C.1	Facebook-Profile linksextremistischer Gruppierungen in Sachsen . . . . .	101
C.2	Facebook-Profile rechtsextremistischer Gruppierungen in Sachsen . . . . .	101

## IV. Abkürzungsverzeichnis

AG .....	Arbeitsgemeinschaft, Seite 21
API .....	Application Programming Interface, Seite 12
App .....	Application, Seite 2
BBS .....	Bulletin Board System, Seite 8
bzw. ....	beziehungsweise, Seite 17
CSV .....	Comma Separated Values, Seite 38
E-Mail .....	Electronic Mail, Seite 11
f. ....	folgende [Seite], Seite 8
ff. ....	folgende [Seiten], Seite 5
FOAF .....	Friend Of A Friend, Seite 38
ID .....	Identification [Number], Seite 16
IRC .....	Internet Relay Chat, Seite 8
ISO .....	International Standardization Organization, Seite 39
Kripo .....	Kriminalpolizei, Seite 21
LfV .....	Landesamt für Verfassungsschutz, Seite 40
Mio. ....	Million[en], Seite 11
Mrd. ....	Milliarde[n], Seite 11
n/a .....	not available, Seite 54
NLP .....	Natural Language Processing, Seite 31
PCA .....	Principal Component Analysis, Seite 25
RDF .....	Resource Description Framework, Seite 38
SemSNA .....	Semantic Social Network Analysis, Seite 38
SIOC .....	Semantically-Interlinked Online Communities, Seite 38
StPO .....	Strafprozessordnung, Seite 21
TKÜ .....	Telekommunikationsüberwachung, Seite 21
usw. ....	und so weiter, Seite 7
vgl. ....	vergleiche, Seite 6
XML .....	Extensible Markup Language, Seite 39
XSD .....	XML Schema Definition, Seite 44
XSLT .....	Extensible Stylesheet Language Transformation, Seite 39



# 1 Einleitung

„I do not want to live in a world  
where everything I do and say is recorded.“  
— EDWARD SNOWDEN [Gua13]

Die scheinbare Anonymität und persönliche Distanz in sozialen Online-Netzwerken begünstigt, dass auch extreme gesellschaftliche und politische Positionen freier geäußert und ausgelebt werden als im persönlichen Diskurs. Die aktuellen Debatten um *Fake News*, *Social Bots* und Wahlmanipulation im Internet zeigen die Relevanz, aber auch Brisanz dieses Themas. Die extremen Positionen und die starke Vernetzung der beteiligten Personen bei gleichzeitiger Entstehung sogenannter Filterblasen sowie gruppendynamische Prozesse können dazu führen, dass soziale Online-Netzwerke zur Planung, Vorbereitung und Durchführung von Straftaten sowohl im virtuellen Raum als auch im physischen Leben genutzt werden. Es ist zu erwarten, dass der Missbrauch sozialer Online-Netzwerke auch für illegale Zwecke weiter zunehmen wird. [ROR16] [Bun16] [Mar16]

## 1.1 Motivation

Gemäß dem Polizeigesetz des Freistaates Sachsen umfassen die Aufgaben der Polizei insbesondere [Sac13]

- den Schutz der freiheitlich demokratischen Grundordnung und Gewährleistung der ungehinderten Ausübung der Grundrechte und staatsbürgerlichen Rechte,
- die Verhinderung und vorbeugende Bekämpfung von Straftaten sowie
- das Treffen von Vorbereitungen, um künftige Gefahren abwehren zu können.

Unter den beschriebenen Bedingungen spielt die Überwachung sozialer Online-Netzwerke eine zunehmend wichtige Rolle, um die polizeilichen Aufgaben hinsichtlich Kriminalprävention und Gefahrenabwehr sowie für Ermittlungsverfahren bei Straftaten auch im virtuellen Raum effektiv und effizient wahrnehmen zu können. Gleichzeitig machen das enorme Volumen und die Komplexität der Daten eine manuelle Bearbeitung nahezu unmöglich, so dass automatisierte Verfahren für die Verarbeitung dieser Daten gefunden werden müssen. Hierfür müssen Mechanismen entwickelt werden, die Personen oder Personengruppen, die zukünftig Straftaten begehen könnten – sogenannte Gefährder –, in sozialen Online-Netzwerken sicher erkennen können (siehe dazu Abschnitt 3.1). [SL13, SHGL15, SBHL17]

In den vergangenen Jahren wurden zahlreiche Studien veröffentlicht, die sich mit der Prognose von Ereignissen und Entwicklungen in sozialen Online-Netzwerken beschäftigen. Verbreitete Themen sind Vorhersagen über den Erfolg von Marketingaktionen und Kinofilmen, die Ausbreitung von Informationen, den Ausgang von Wahlen und Indikatoren für wirtschaftliche Entwicklung. [YK12] Darüber hinaus gibt es verschiedene Untersuchungen zur Detektion von Nutzern mit abweichendem Verhalten in sozialen Online-Netzwerken (*Social Spammer*, *Fake-Profile*) [LCW10, WKA16, KS16] oder Klassifizierung von Nutzern hinsichtlich verschiedener Aspekte wie Nutzung von *FACEBOOK-Apps*, politische Einstellung oder ethnische Zugehörigkeit [PP11, RN16]. Alle Beispiele haben gemeinsam, dass sie entweder auf thematisch sehr engen Datensammlungen basieren oder sich auf theoretische Vorschläge für Vorgehensmodelle und Forschungsansätze beschränken. Damit lassen sich deren Erkenntnisse und Aussagen nur schlecht auf andere Domänen übertragen, was verschärft auf so spezielle Anwendungsgebiete wie das der vorliegenden Arbeit zutrifft.

Auch im Kontext des sogenannten *Predictive Policing* (deutsch: vorausschauende Polizeiarbeit) gibt es einige Arbeiten, die sich jedoch überwiegend mit der Anwendung von Methoden der Analyse sozialer Netzwerke (im konventionellen Sinne) und des *Data Mining* auf existierende Daten aus polizeilichen und anderen behördlichen Datenbanken beschäftigen [WCDO09, DS10, TG16]. Andere Untersuchungen im Umfeld sozialer Online-Netzwerke stellen fest, dass es bewährte und gut funktionierende Methoden, Algorithmen und Werkzeuge aus dem *Data Mining* und *Maschinellen Lernen* gibt, die aktuelle Verfügbarkeit und Qualität der relevanten Daten als entscheidende Faktoren jedoch nicht ausreichend für gute und verlässliche Vorhersagen sind [BM11, PB16].

Bisher beschäftigt sich die sogenannte vorhersagebasierte Polizeiarbeit im Wesentlichen mit der Prognose räumlicher und zeitlicher Schwerpunkte des Auftretens bestimmter Deliktarten; meist handelt es sich um Wohnungseinbruchdiebstahl. Diese Prognosen werden dann meist für die Planung von Stärke sowie Ort und Zeit der Präsenz von Einsatzkräften herangezogen. Die Prognosen werden durch Auswertung polizeilicher Falldaten und teilweise zusätzlich öffentlich zugänglicher infra- und soziostruktureller sowie sozioökonomischer Daten mittels Verfahren des *Data Mining* und *Maschinellen Lernens* erstellt. International gibt es Bestrebungen, mit dem sogenannten *Predictive Profiling* personenbezogene Vorhersagen über Täter- und Opferschaft auf Basis von Informationen aus verschiedenen amtlichen Datenbanken (beispielsweise über Vorstrafen, Wohnort und Umfeld) zu generieren. Problematisch ist in allen Fällen, dass für die Prognosen naturgemäß nur die Daten des Hellfeldes der Kriminalität herangezogen werden können und es bisher keine belastbaren Ergebnisse über die Wirkung von *Predictive Policing* gibt. [Egb17]

Unter Berücksichtigung der aktuellen Herausforderungen der Polizeiarbeit und der beschriebenen methodischen Defizite wird die Notwendigkeit deutlich, neue Ansätze für die Auswertung von Daten aus sozialen Online-Netzwerken zu entwickeln. Für die Ent-



wicklung von Modellen für das *Maschinelle Lernen* werden standardisierte Referenzdaten benötigt, welche die Bewertung und den Vergleich dieser Modelle und deren Performance ermöglichen. Für den vorliegend betrachteten Fall der Detektion von Gefährdern in sozialen Online-Netzwerken ist ein solcher Datensatz bisher nicht verfügbar.

## 1.2 Zielstellung

Ziel der vorliegenden Arbeit ist die Entwicklung eines Datensatzes als Goldstandard für die Konstruktion, das Training und die Evaluation von Klassifizierungsalgorithmen, um Gefährder in sozialen Netzwerken zu detektieren (siehe Kapitel 3). Wichtige Fragestellungen dieser Arbeit sind Umfang, Repräsentativität und Vollständigkeit der Referenzdaten. In diesem Zusammenhang sollen relevante Trainingsbeispiele und *Features* für den Goldstandard gefunden sowie ein Vorschlag für deren Bereitstellung erarbeitet werden, um die Voraussetzungen für eine möglichst breite und universelle Nutzung zu schaffen (siehe Abschnitt 4.1). Die hervorragende Eignung von FACEBOOK als Forschungsobjekt im Allgemeinen (siehe Abschnitt 2.4) und die gute Sichtbarkeit der Eigenschaften und Aktivitäten von FACEBOOK *Pages* im Speziellen (siehe Abschnitt 2.4.2) waren ausschlaggebend dafür, dass sich die vorliegende Arbeit für die Entwicklung des Goldstandards zunächst auf FACEBOOK *Pages* konzentriert.

## 1.3 Aufbau der Arbeit

Die Arbeit gliedert sich in drei thematische Bereiche. Das Kapitel 2 beschäftigt sich zunächst mit der Analyse sozialer Netzwerke als Gegenstand der vorliegenden Arbeit. Neben der *Graphentheorie* als Konzept für die Beschreibung sozialer Netzwerke und deren Beziehungen werden auch Eigenschaften von *Social Media* und sozialen Online-Netzwerken im Allgemeinen und von FACEBOOK im Speziellen betrachtet. Im anschließenden Kapitel 3 werden die Grundlagen für die Detektion von Gefährdern im Kontext von *Maschinellern Lernen* und *Social Media* erläutert. Diese beiden Kapitel bilden das Fundament für die Entwicklung des *Goldstandards* mit Überlegungen zu Anforderungen und Umsetzung in Kapitel 4. Das abschließende Kapitel 5 fasst die Ergebnisse der vorliegenden Arbeit zusammen und gibt einen Ausblick auf offene Fragestellungen und mögliche zukünftige Weiterentwicklungen.



## 2 Analyse sozialer Netzwerke

Das Konzept sozialer Netzwerke als explizite Beschreibung zwischenmenschlicher Strukturen stammt bereits aus der ersten Hälfte des vorigen Jahrhunderts. Die deutschen Soziologen VIERKANDT und VON WIESE verwendeten in den 1920er Jahren erstmals die Terminologie von Punkten, Linien und Verbindungen, um soziale Beziehungen zwischen Akteuren zu beschreiben. MORENO präsentierte 1934 seine Forschung zum Netzwerkcharakter sozialer Beziehungen und nannte seinen Ansatz *Soziometrie*. Er entwickelte die Idee, soziale Strukturen als Diagramme aus Punkten und Linien – sogenannte *Soziogramme* – darzustellen [Mor34, S. 21ff.]. Die *Soziometrie* wurde schnell zu einem wichtigen Forschungsgebiet der Anthropologie und Soziologie, aber auch der Psychologie und Pädagogik. Schwerpunkte waren unter anderem die Untersuchung von Gruppenstrukturen und gruppendynamischer Prozesse, wie beispielsweise die Bildung sogenannter *Cliquen*. [Sco11]

### 2.1 Exkurs: Graphentheorie

Der vorherrschende Ansatz zur Analyse sozialer Netzwerke ist das mathematische Konzept der *Graphentheorie*. Ein *Graph* besteht aus einer Menge von *Knoten* (Akteure im sozialen Netzwerk) und einer Menge von *Kanten* (Beziehungen zwischen den Akteuren). Kanten repräsentieren paarweise Verbindungen zwischen Knoten und können gerichtet und ungerichtet sein. Zusätzlich können Knoten und Kanten mit *Gewichten* versehen werden, um die Stärke der Verbindungen zu beschreiben. Wenn zwei Knoten durch eine Kante verbunden sind, werden sie als *benachbart* bezeichnet. [BH83, Sco11]

Es gibt verschiedene Metriken, um die Eigenschaften eines Graphen zu beschreiben. Zwei wichtige Merkmale sind neben der Anzahl der Knoten und Kanten die *Dichte* des Graphen und die *Zentralität* von Knoten. Die Dichte oder auch Kantendichte eines Graphen gibt an, wie groß das Verhältnis der Anzahl vorhandener Kanten zur Anzahl theoretisch maximal möglicher Kanten ist. Die Zentralität ist ein Maß für die Wichtigkeit eines Knotens im Netzwerk und kann aus verschiedenen Perspektiven bestimmt werden [TL10, S. 13ff.] [Sco11]:

**Gradzentralität** Die Bedeutung eines Knotens wird durch die Anzahl der unmittelbar benachbarten Knoten bestimmt. Je mehr Nachbarn ein Knoten hat, desto höher ist sein *Grad* (englisch: *degree*) und umso wichtiger ist der Knoten im Netzwerk. Bei gerichteten Graphen werden eingehende (*in-degree*) und ausgehende (*out-degree*) Verbindungen unterschieden. In den meisten sozialen Netzwerken (sowohl in der realen als auch in der Online-Welt) sind die Knotengrade potenzverteilt, das heißt sehr wenige Knoten

vereinen eine sehr große Anzahl von Verbindungen auf sich, wohingegen die meisten Knoten nur vergleichsweise wenige Verbindungen haben.

**Nähezentralität** Zentrale Knoten haben eine große Nähe – also geringe Distanz – zu anderen Knoten im Netzwerk. Üblicherweise werden Knoten mit höherer Zentralität als wichtiger angenommen, so dass die Nähezentralität durch den reziproken Wert der durchschnittlichen Pfadlänge eines Knotens zu allen anderen Knoten des Netzwerks angegeben wird. In sozialen Netzwerken lässt sich die durchschnittliche Pfadlänge als Maß interpretieren, wie schnell eine Information von einem Knoten ausgehend das gesamte Netzwerk erreichen kann.

**Zwischenzentralität** Die Zwischenzentralität eines Knotens drückt aus, wie häufig dieser Knoten Teil der kürzesten Pfade zwischen den Knotenpaaren eines Netzwerks ist. Über wichtige Knoten laufen die meisten Verbindungen, so dass diese eine Schlüsselposition bei der Kommunikation innerhalb eines sozialen Netzwerks einnehmen. Knoten mit einer hohen Zwischenzentralität fungieren oft als Brücke zwischen Subnetzen und verbinden andernfalls getrennte Teile eines Netzwerks miteinander.

**Eigenvektorzentralität** Die Bedeutung eines Knotens im Sinne der Eigenvektorzentralität wird durch die Wichtigkeit seiner benachbarten Knoten definiert. Je mehr und wichtigere Nachbarn ein Knoten besitzt, desto höher ist seine Eigenvektorzentralität und damit seine eigene Bedeutung. Beispielsweise kann der GOOGLE PAGERANK als eine Variante der Eigenvektorzentralität betrachtet werden, der neben anderen Merkmalen Teil der Rankingfunktion ist, welche die Relevanz von Suchergebnissen der GOOGLE SUCHE bewertet (vgl. [BP98]).

Es wird deutlich, dass die Bedeutung eines Knotens im Netzwerk nicht durch seine individuellen Attribute bestimmt wird, sondern von seiner Position und Verknüpfung innerhalb des Netzwerks abhängt. Verschiedene Untersuchungen haben jedoch auch gezeigt, dass es Korrelationen zwischen den Attributen von Akteuren in einem sozialen Netzwerk und deren Verhalten gibt. Die Forschung auf dem Gebiet der *Social Network Analysis* führt das darauf zurück, dass Akteure mit gemeinsamen Attributen auch ähnliche Positionen in sozialen Netzwerken besetzen und diese deshalb ähnlich agieren respektive deren Handeln ähnliche Auswirkungen hat. [BMBL09, MW11, LWC14]

Das Bestimmen der oben beschriebenen Maßzahlen von Graphen ist schon für lokal begrenzte soziale Netzwerke mit einem sehr großen Aufwand verbunden und in Online-Netzwerken auf Grund der schieren Masse an Daten nahezu unmöglich. Schon das Erheben der Daten innerhalb einer vernünftigen Zeitspanne ist kaum denkbar, in jedem Fall aber stoßen die bekannten Algorithmen für das Berechnen der Metriken im globalen Maßstab auf Grund der Komplexität sehr schnell an ihre technischen Grenzen. Es gibt verschiedene Strategien, die Komplexität zu reduzieren, um diesen Skalierungsproblemen zu begegnen, beispielsweise das Analysieren von Stichproben oder das Fokussie-

ren auf lokale Teilgraphen, die jeweils nur einen kleinen Teil des gesamten Netzwerks abbilden. [PKVS12]

## 2.2 Beziehungen in sozialen Netzwerken

Verbindungen zwischen den Akteuren eines sozialen Netzwerks werden nicht nur durch explizite soziale Beziehungen, wie beispielsweise Verwandtschaft und Freundschaft, gebildet, sondern können auch durch Interaktionen sowie immateriellen und materiellen Austausch zwischen Akteuren entstehen. Darüber hinaus können Verbindungen auch implizit sein, wenn Akteure im Netzwerk beispielsweise gemeinsame demografische Merkmale oder Interessen teilen. BORGATTI et al. identifizieren vier Kategorien von Beziehungen in sozialen Netzwerken [BMBL09, MW11]:

**Gemeinsamkeiten** Indirekte Beziehungen zwischen Akteuren können betrachtet werden, wenn diese gemeinsame Attribute wie Wohnort, demografische Merkmale (Geschlecht, Alter usw.), Interessen und Gruppenzugehörigkeiten (gleicher Verein, gleiche Clubs, gleiche Veranstaltungen usw.) teilen.

**Soziale Beziehungen** Verwandtschaft, Freundschaft und andere Rollenbeziehungen (Vorgesetzter, Kommilitonin usw.) definieren explizite Verbindungen in sozialen Netzwerken, ebenso wie emotionale (*jemanden mögen*) und kognitive (*jemanden kennen*) Beziehungen.

**Interaktionen** Verbindungen zwischen Akteuren können durch Interaktionen wie Sprechen, Beraten oder Helfen entstehen. Diese verhaltensbasierten Bindungen treten meist im Kontext sozialer Beziehungen auf.

**Flüsse** Verbindungen können auch auf dem Austausch oder Transfer zwischen Knoten basieren. Dabei fließen unter anderem Informationen, Meinungen oder Ressourcen durch ein soziales Netzwerk. Ebenso wie Interaktionen treten Flüsse häufig innerhalb anderer sozialer Beziehungen auf.

Übertragen auf Online-Netzwerke bedeutet dies, dass neben expliziten sozialen Beziehungen (*Freund, Kontakt, Follower* usw.) auch Interaktionen wie das *Bewerten, Teilen* und *Kommentieren* von Objekten in sozialen Netzwerken Knoten miteinander verbinden können. Das impliziert gleichzeitig, dass es verschiedene Typen von Verbindungen zwischen Knoten eines sozialen Netzwerks geben muss, um die unterschiedliche Herkunft und Interpretation dieser Verbindungen zu berücksichtigen. Darüber hinaus müssen die Knoten eines Online-Netzwerks ebenfalls typisiert sein, um die Verbindungen zwischen heterogenen Akteuren – wie beispielsweise zwischen Personen oder Organisationen und Veranstaltungen oder Gruppen – abbilden zu können.

Um die beschriebenen Anforderungen an die Modellierung sozialer Netzwerke besser abbilden zu können, schlugen FARARO und DOREIAN 1984 zunächst Graphen mit bis zu drei verschiedenen Knotentypen vor (*two-mode* und *three-mode networks*) [FD84]. Diese wurden später zu *Multi-Mode*-Netzwerken mit beliebig vielen Knotentypen verallgemeinert und um *Multi-Dimensionalität* erweitert, welche auch Verbindungen unterschiedlichen Typs zwischen den Knoten berücksichtigt. [TL10, S. 68ff.] [MW11, MBW13]

## 2.3 Soziale Online-Netzwerke

Erste Vorläufer sozialer Online-Netzwerke erlangten bereits in den 1980er Jahren mit dem *Usenet* für Diskussionsforen (*Newsgroups*), *Internet Relay Chat (IRC)* als Chat-System und *Bulletin Board Systems (BBS)* zum Austausch von Nachrichten und Dateien in akademischen und computeraffinen Gruppen vergleichsweise große Popularität. [EPK<sup>+</sup>11] [TAH<sup>+</sup>16, S. 13] Die sprunghafte Entwicklung der technischen Möglichkeiten, aber auch der Medienkompetenz der Nutzer führte Anfang der 2000er Jahre zu einer zunehmenden Verbreitung interaktiver und kollaborativer Anwendungen im Internet. Diese neue Evolutionsstufe des Internets ist mit Schlagwörtern wie *Web 2.0*, *Social Media* und *User-generated Content* verbunden. Bekannte Vertreter der sogenannten *Social Software* sind unter anderem *Wikis*, *Weblogs* und *Instant Messenger* sowie *Plattformen* zum Teilen von Fotos oder Musik. Auf die ersten, meist geschlossenen, *Online-Communities* der 1990er Jahre folgten die heutigen *sozialen Online-Netzwerke*, die eine ganz neue Sichtweise auf die Analyse sozialer Netzwerke eröffneten. [BE07, EPK<sup>+</sup>11, HKP12]

Ziele der Analyse sozialer Online-Netzwerke sind unter anderem das Finden von Gruppenstrukturen (*Communities*), das Identifizieren von Schlüsselpositionen und Rollen (Meinungsführer, Multiplikatoren, ...), das Aufdecken verborgener Beziehungen sowie das Extrahieren von Meinungen und Themen (*Opinion* und *Topic Mining*). [Atz12, GL12]

### 2.3.1 Exkurs: Social Media

Der Begriff *Social Media* beschreibt moderne Formen der interaktiven und kollaborativen Nutzung digitaler Technologien zur Kommunikation sowie zum Erstellen und Austauschen von Informationen und grenzt diese von traditionellen, unidirektionalen Kanälen (Buch, Zeitung, Radio, Fernsehen) ab, bei denen die Rollen von Produzent und Rezipient der Kommunikation klar getrennt sind. Mit *Social Media* werden zum einen das Konzept der gemeinsamen Mediennutzung und zum anderen die konkreten technischen Plattformen und Anwendungen (*Social Software*) bezeichnet. [TL10, S. 1f.]

KIETZMANN et al. haben sieben Funktionsbereiche identifiziert, welche die verschiedenartigen Facetten von *Social Media* repräsentieren. Diese Aspekte sind bei unterschiedlichen Typen von *Social-Media-Plattformen* verschieden stark ausgeprägt und es müssen auch nicht zwingend alle sieben Funktionsbereiche vorhanden sein [KHMS11]:

1. **Identity** – Darstellen der eigenen Online-Persönlichkeit,
2. **Conversations** – Kommunizieren mit anderen Nutzern,
3. **Sharing** – Austauschen, Verteilen und Empfangen von Inhalten,
4. **Presence** – Anwesenheit zu einer bestimmten Zeit oder an einem bestimmten Ort,
5. **Relationships** – Verbindungen unterschiedlicher Qualität mit anderen Nutzern,
6. **Reputation** – Erkennen der Reputation (Vertrauen, Ruf, Ansehen) und Bewerten von Personen und Inhalten,
7. **Groups** – Bilden von Gemeinschaften anhand sozialer Beziehungen (beispielsweise Familie und Freunde) oder anderer Kriterien (Interessen usw.).

Je nach Fokus auf einen oder mehrere Funktionsbereiche können verschiedene Formen von *Social Media* unterschieden werden (in Anlehnung an [MGN12] [GL12]):

- **Wikis** – gemeinsames Erstellen von Inhalten
- **Foren** – Diskussion und Ratgeber
- **Weblogs** – Publizieren von Online-Tagebüchern
- **Microblogs** – Teilen von Kurznachrichten
- **Social News** – Teilen von Links und Artikeln
- **Social Gaming** – Gemeinsame Online-Spiele
- **Media Sharing** – Teilen von Foto, Video, Musik, Livestreams, Podcasts
- **Chats und Messenger** – Kommunikation zwischen Nutzern oder in Gruppen
- **Location based Services** – Austausch und Aktivitäten auf Basis des gleichen geografischen Standorts
- **Bewertungs- und Verbraucherportale** – Bewertung von Produkten und Dienstleistungen, Beantworten von Nutzerfragen
- **Soziale Online-Netzwerke** – Vernetzung zwischen Nutzern mit gemeinsamen Beziehungen oder Interessen

Zwischen den verschiedenen Formen von *Social Media* bestehen vielfach Überschneidungen und eine exakte Abgrenzung ist oftmals nicht möglich. Auch wenn beispielsweise bei sozialen Online-Netzwerken der Vernetzungsaspekt im Vordergrund steht, integrieren diese darüber hinaus eine Vielzahl anderer *Social-Media*-Funktionen. Auf der anderen Seite ist es so, dass auch die meisten der anderen *Social-Media*-Anwendungen zumindest rudimentäre Funktionen eines sozialen Netzwerks – wie Nutzerprofile und Vernetzung – integrieren, so dass beispielsweise auch der *Microblogging*-Dienst TWITTER oft als soziales Online-Netzwerk bezeichnet wird. Auch Online-Plattformen, die

nicht den originären *Social-Media*-Anwendungen zugerechnet werden, bieten ihren Nutzern häufig entsprechende Funktionen, um ihre Angebote attraktiver zu machen (beispielsweise EBAY, AMAZON).

### 2.3.2 Merkmale und Funktionen

Soziale Online-Netzwerke sind *Online-Communities*, die nach verschiedenen Kriterien, wie beispielsweise Zielgruppe (Privatpersonen, Geschäftsleute, Wissenschaftler), lokaler Fokus (Nachbarschaft, Region, Land) oder thematische Spezialisierung (Schulfreunde finden, Fotos teilen) differenziert werden können. Auf Basis von sozialen Online-Netzwerken als technische Plattform können einerseits soziale Netzwerke aus ausgewählten oder allen Bereichen des realen Lebens in die Online-Welt übertragen und andererseits bisher nicht (real) existierende soziale Netzwerke aufgebaut werden. Online-Netzwerke bieten dabei auf Grund der technischen Möglichkeiten und ihrer potenziell riesigen Reichweite eine ganz andere Qualität sozialer Interaktionen bis hin zur Möglichkeit, eine komplett eigene *Online-Identität* einer realen oder erdachten Person zu entwerfen. [BE07, Kos14]

Trotz der vielen verschiedenen Facetten bieten die meisten sozialen Online-Netzwerke einen vergleichbaren Umfang an Basisfunktionen, die mehr oder weniger stark ausgeprägt, adaptiert oder durch spezielle Merkmale erweitert sein können (vgl. [HKP12], [TAH<sup>+</sup>16, S. 20ff.]):

- **Persönliches Profil** mit Informationen zur Selbstdarstellung (Foto, persönliche und demografische Angaben, Interessen, Aktivitäten),
- **Vernetzung** mit anderen Nutzern (Aufbauen von Verbindungen im Netzwerk, Verwalten und Kategorisieren von Kontakten),
- Erstellen von privaten und öffentlichen **Gruppen** innerhalb des Netzwerks,
- **Veröffentlichen von Inhalten** (Text, Link, Bild, Video, Audio, bspw. Statusmeldungen, Erlebnisse),
- **Check-in** an bestimmten Orten (Geschäft, Restaurant, Sehenswürdigkeit, Stadt),
- **Teilen, Bewerten und Kommentieren** von Inhalten,
- Zentrale **Übersicht der Aktivitäten** im persönlichen Netzwerk,
- **Nachrichtensystem** zur Kommunikation zwischen einzelnen oder mehreren Nutzern (Austausch von Nachrichten und Dateien),
- **Applikationen** zur Interaktion im Netzwerk (Gewinnspiele, Umfragen, Chat-Bots, Online-Kauf, Online-Spiele),
- **Suche** nach Nutzern und Inhalten,
- **Benachrichtigungen** über Ereignisse im Netzwerk,
- **Einstellungen** zu Privatsphäre und Datenschutz **sowie Konfiguration** des Nutzerkontos.



Die meisten Aktivitäten und Interaktionen in sozialen Online-Netzwerken finden nahezu in Echtzeit statt. Sobald ein Nutzer eines Online-Netzwerks einen Status oder andere Inhalte publiziert, teilt, bewertet oder kommentiert, fließen diese Aktionen unmittelbar in den Ereignisstrom der mit ihm verbundenen Nutzer (und unter Umständen auch deren Netzwerk-Nachbarn) ein und die potenziellen Empfänger dieser Kommunikation werden aktiv per E-Mail oder über Push-Mechanismen auf Mobilgeräten darüber informiert. [SRML09, Fac17b]

Sobald ein Nutzer einen gewissen Grad der Vernetzung erreicht hat, wird die Menge an Informationen aus dem persönlichen Netzwerk schnell unübersichtlich. In vielen Online-Netzwerken wird der Nachrichtenstrom deshalb durch den Anbieter der Plattform entsprechend einer antizipierten Relevanz (in Abhängigkeit von Profil, Verbindungen und Aktivitäten eines Nutzers) gefiltert und sortiert, wobei dies durch den betreffenden Nutzer in bestimmten Grenzen gesteuert und beeinflusst werden kann. Darüber hinaus nutzen viele Betreiber sozialer Online-Netzwerke die umfassenden Informationen über die Interessen ihrer Nutzer für die Monetarisierung ihres Geschäftsmodells, indem maßgeschneiderte Werbung und bezahlte redaktionelle Beiträge in den Nachrichtenstrom integriert werden. [SRML09, Fac17b]

Die Echtzeitkomponente sozialer Online-Netzwerke führt auch dazu, dass sämtliche Vorgänge im Vergleich zu sozialen Netzwerken in der realen Welt (beispielsweise der Informationsfluss durch das Netzwerk) deutlich schneller ablaufen und eine enorme Dynamik entwickeln können. Die Funktion der Zeit als Moderator (Entschleuniger) von Kommunikation ist in sozialen Online-Netzwerken weniger wirksam; die zeitliche Verzögerung zwischen Aktion und Reaktion (zum Innehalten und zur Reflexion des Gesagten oder Geschriebenen) wird durch die synchrone Benachrichtigung und einfache Möglichkeiten zur unmittelbaren Reaktion stark reduziert. [Cen10]

Gleichzeitig sind diese Intensität und Dynamik der Kommunikation in sozialen Online-Netzwerken auch ein Grund für die riesigen Datenmengen, die Wirtschaft und Forschung wertvolle Informationen über die Nutzer (und teilweise Nicht-Nutzer) dieser Online-Netzwerke liefern können und die auch Gegenstand der vorliegenden Arbeit sind.

## 2.4 Forschungsobjekt Facebook

FACEBOOK ist mit derzeit täglich mehr als 1,3 Mrd. und monatlich über 2 Mrd. aktiven Nutzern das weltweit größte soziale Online-Netzwerk (Stand Juli 2017, vgl. [Fac17c, Sta17]). Mit Stand vom August 2017 leben ca. 7,56 Mrd. Menschen auf der Erde (vgl. [Deu17c]) – damit ist mehr als ein Viertel der aktuellen Weltbevölkerung mindestens einmal monatlich auf FACEBOOK aktiv. In Deutschland nutzen ca. 30 Mio. Menschen monatlich und 23 Mio. Menschen täglich FACEBOOK (Stand Juni 2017, vgl. [Fac17a]). FACEBOOK zählt nach GOOGLE und YOUTUBE sowohl in Deutschland als auch weltweit

zu den meist besuchten Websites überhaupt (vgl. [Ale17a, Ale17b]). Neben den persönlichen Profilen sind auch 70 Mio. Unternehmen mit einer eigenen Seite in FACEBOOK vertreten (Stand Juli 2017, vgl. [Fac17c]).

Schon auf Grund der beschriebenen Marktpenetration in Bevölkerung, Wirtschaft und auch Verwaltung sind FACEBOOK und die Daten seiner Nutzer als Forschungsgegenstand sehr attraktiv. KOSINSKI identifiziert darüber hinaus verschiedene Vorteile von sozialen Online-Netzwerken im Allgemeinen und FACEBOOK im Speziellen, die für deren ausgezeichnete Eignung in der Forschung sprechen [Kos14, S. 6ff.]:

**Große und repräsentative Stichproben** Obwohl es Bevölkerungsgruppen gibt, die aus verschiedenen Gründen auf FACEBOOK nicht oder wenig präsent sind, kann davon ausgegangen werden, dass es weltweit keinen größeren und repräsentativeren Pool an Teilnehmern für Studien gibt.

**Wenig Artefakte** Ein soziales Online-Netzwerk ermöglicht das Beobachten von tatsächlichem und natürlichem Verhalten der Nutzer in einer natürlichen Umgebung (*natürlich* in dem Sinne, dass eine Situation nicht für eine Befragung oder Beobachtung künstlich erzeugt wird und dadurch das Verhalten beeinflussen kann). FACEBOOK selbst geht auch davon aus, dass es sich nur bei einem vergleichsweise geringen Anteil von ca. 7 % der monatlich aktiven Nutzer um Accounts handelt, die keinen realen Nutzer repräsentieren (Duplikate, Falschzuordnungen, *Fake-Accounts*, vgl. [Fac17d]).

**Valide Daten** Die beobachteten Aktivitäten von Nutzern werden in Online-Netzwerken über einen langen Zeitraum aufgezeichnet und nicht explizit durch die Teilnehmer für eine Untersuchung erfasst, womit versehentliche Fehler oder absichtliche Falschaussagen unwahrscheinlicher sind.

**Größere Offenheit** Nutzer auf FACEBOOK und anderen sozialen Online-Netzwerken zeigen im Vergleich zu konventionellen Befragungen eine größere Bereitschaft, persönliche Informationen zu teilen (dieser Aspekt könnte sich in Hinblick auf das aktuelle Umdenken hinsichtlich Datenschutz und Privatsphäre in sozialen Online-Netzwerken in Zukunft deutlich abschwächen und müsste dann möglicherweise neu bewertet werden).

**Bequemlichkeit** Studien in sozialen Online-Netzwerken bedeuten für die Probanden keinen zusätzlichen Aufwand, da sämtliche Daten und Informationen bereits „erfasst“ respektive aufgezeichnet sind.

**Einfacher Zugriff** Daten über Nutzer und Aktivitäten aus FACEBOOK können im Rahmen der Nutzungsbedingungen für Forschungszwecke relativ einfach über die FACEBOOK *Graph API* abgefragt werden (vgl. [Fac17e]).

**Archivcharakter** Die Daten aus sozialen Online-Netzwerken repräsentieren eine Art Archiv der Netzwerkaktivitäten eines Nutzers. In der Vergangenheit geknüpfte Verbindungen zu Personen, Unternehmen und Gruppen bleiben in vielen Fällen erhalten, auch wenn sich Einstellungen, Interessen oder Freundschaften ändern.

Vor allem die große Verbreitung, die gute Verfügbarkeit valider und repräsentativer Daten sowie der einfache technische Zugriff auf den *Social Graph* machen FACEBOOK für die Forschung sehr attraktiv. Aus diesen Gründen liegt auch der Schwerpunkt dieser Arbeit auf der Auswertung von Daten aus FACEBOOK.

## 2.4.1 Facebooks Social Graph

Die Datenstrukturen von FACEBOOK orientieren sich technisch stark am bereits vorgestellten Konzept eines *Graphen* und werden von FACEBOOK auch als *Social Graph* bezeichnet. Der FACEBOOK *Social Graph* kennt drei Elemente: *Nodes*, *Edges* und *Fields*<sup>1</sup>. [Fac17e]

Stark vereinfacht repräsentieren *Nodes* diejenigen Entitäten auf FACEBOOK (*Knoten* des *Graphen*), mit denen ein Nutzer interagieren kann, wobei auch ein Nutzer einen *Node* in diesem Sinne darstellt. *Edges* in FACEBOOK sind entsprechend der Graphentheorie die Verbindungen (*Kanten*) zwischen den *Nodes*. *Fields* wiederum beschreiben die Eigenschaften (*Attribute*) von *Nodes*. *Fields* können elementare und zusammengesetzte Datentypen – wie *ganze Zahlen* und *Zeichenketten* – oder Datenstrukturen – wie *Listen* und *Felder* – beinhalten. [Fac17b, Fac17e]

FACEBOOK unterscheidet verschiedene Typen sogenannter *Profile Nodes* [Fac17e]:

- **Application** – Anwendungen, Spiele und *soziale Plugins*, um mit anderen Nutzern zu interagieren,
- **Event** – Veranstaltungen in der realen Welt,
- **Group** – Öffentlicher oder privater Raum zur Kommunikation mit Nutzern über gemeinsame Interessen,
- **Page** – Profil für Marken, Unternehmen, Organisationen und Personen des öffentlichen Lebens,
- **User** – Profil für natürliche Personen.

Neben den *Profile Nodes* gibt es weitere *Node*-Typen, welche die verschiedenen Entitäten, aber auch funktionale Elemente von FACEBOOK repräsentieren.

<sup>1</sup> Begriffe, die technische Konzepte der FACEBOOK *Graph API* beschreiben, werden im Rahmen dieser Arbeit nur übersetzt, sofern es der besseren Lesbarkeit oder dem Verständnis dient. Alle Angaben beziehen sich auf die mit Stand vom August 2017 aktuelle Version 2.10 der FACEBOOK *Graph API*. Als *deprecated* (veraltet) markierte Elemente werden in Übersichten mit aufgeführt, aber für die weitere Arbeit nicht berücksichtigt.

Im Kontext dieser Arbeit relevante *Node*-Typen sind unter anderem [Fac17e]:

- **Album** – Fotoalbum zum Verwalten von Fotos,
- **Comment** – Kommentar zu Beiträgen oder Interaktionen im Netzwerk,
- **Photo** – Foto, das im Netzwerk geteilt oder in Fotoalben eingefügt werden kann,
- **Place** – Ortsangabe zum Teilen des Aufenthaltsortes,
- **Post** – Beitrag, welcher in der Chronik eines *Users*, einer *Page*, einer *Group* oder einer *Application* veröffentlicht wird,
- **Video** – Video, das im Netzwerk geteilt oder in Alben eingefügt werden kann.

Jeder *Node* besitzt einen definierten Satz an *Fields*, welche die Eigenschaften eines *Nodes* beschreiben. Ein *User Node* beispielsweise hat über 50 *Fields* und wird unter anderem durch folgende Angaben charakterisiert [Fac17e]:

- **id** – Eindeutige Identifikationsnummer,
- **birthday** – Geburtstag,
- **gender** – Geschlecht,
- **hometown** – Heimatort,
- **location** – Wohnort,
- **name** – Vor- und Nachname,
- **religion** – Religion.

Jeder *Node* kann über verschiedene Typen von *Edges* mit anderen *Nodes* verbunden sein. Der Typ des oder der verbundenen *Nodes* wird durch den Typ einer *Edge* bestimmt. Ein *User Node* kann unter anderem über folgende *Edges* mit anderen *Nodes* verbunden sein [Fac17e]:

- **feed** – durch den Nutzer veröffentlichte Beiträge, Statusaktualisierungen und Links,
- **friends** – mit dem Nutzer befreundete Nutzer,
- **groups** – Gruppen, welche der Nutzer administriert,
- **likes** – Seiten, welche der Nutzer geliked hat,
- **photos** – Fotos, welche der Nutzer hochgeladen hat oder auf denen er markiert wurde.

Die beschriebenen Elemente bilden nur eine kleine Teilmenge der verfügbaren Objekte ab, um die grundsätzliche Struktur des FACEBOOK *Social Graph* zu verdeutlichen. Bereits dieser Ausschnitt zeigt eindrücklich, dass es sich beim FACEBOOK *Social Graph* um ein *multi-modales*, *multi-dimensionales* und darüber hinaus *multi-temporales* soziales Online-Netzwerk handelt (vgl. Abschnitt 2.2). FACEBOOK kennt darüber hinaus eine große Zahl weiterer Typen für Knoten und Kanten, um Akteure und Beziehungen eines sozialen Netzwerks zu beschreiben und den angebotenen Funktionsumfang der FACEBOOK-Plattform zu realisieren. Die im Kontext der vorliegenden Arbeit besonders relevanten Knoten vom Typ *Page* werden im folgenden Abschnitt genauer betrachtet.

## 2.4.2 Anatomie einer Facebook Page

In FACEBOOK werden persönliche und institutionelle Profile unterschieden. Jeder Nutzer auf FACEBOOK benötigt ein persönliches Konto (*Personal Account*), welches als *User Node* im *Social Graph* abgebildet wird. Jeder Nutzer mit einem persönlichen Konto kann eine sogenannte FACEBOOK *Page* erstellen. Eine *Page* ist ein FACEBOOK-Profil, welches für Marken, Unternehmen, Organisationen und Personen des öffentlichen Lebens angelegt werden kann. [Fac17b]

Bei persönlichen FACEBOOK-Profilen werden *Datenschutz* und *Privatsphäre* umfassend unterstützt und die Sichtbarkeit von Informationen und Aktivitäten für andere Nutzer kann sehr granular beeinflusst werden. Viele Nutzer haben in den letzten Jahren erkannt, dass sie selbst dafür verantwortlich sind, ihre Daten vor Missbrauch zu schützen, und haben ihre Profile so konfiguriert, dass fast alle Informationen und Aktivitäten nur noch für verbundene Freunde und nicht öffentlich sichtbar sind. [RGG16]

Im Unterschied dazu kann die Sichtbarkeit von FACEBOOK *Pages* und deren Aktivitäten nur wenig beeinflusst werden. Hintergrund ist, dass *Pages* primär mit dem Ziel der Öffentlichkeitsarbeit angelegt werden, also Unternehmen, Organisationen und Prominente zu präsentieren und zu vermarkten. Die globale Sichtbarkeit einer *Page* im FACEBOOK-Netzwerk kann nur hinsichtlich des Herkunftslandes und des Alters von Nutzern beschränkt werden, die Sichtbarkeit einzelner Beiträge (*Posts*) in Bezug auf Alter, Geschlecht, Wohnort und Sprache der Nutzer. [Fac17b]

Dies hat den für die Forschung entscheidenden Vorteil, dass bei FACEBOOK-*Pages* im Unterschied zu persönlichen Profilen – im Rahmen der Nutzungsbedingungen – viele relevanten Informationen und Aktivitäten abgefragt und analysiert werden können, ohne dass der Administrator (Betreiber) einer *Page* um Zustimmung gebeten werden oder diese erteilen muss. Daten und Funktionen, die der internen Administration und Auswertung dienen, sind über die FACEBOOK *Graph API* naturgemäß ohne entsprechende Berechtigung auch für *Page Nodes* nicht verfügbar (beispielsweise die Liste der durch die *Page* geblockten Nutzer und Seiten).

**Fields** Die Eigenschaften eines *Page Node* werden durch über 120 *Fields* (Attribute) beschrieben. Einige der *Fields* sind nur für bestimmte Arten von Seiten (beispielsweise Restaurants) oder spezielle Anwendungsfälle (beispielsweise eine mobile *App*) anwendbar. Auf einige *Fields* kann ohne zusätzliche Berechtigung (beispielsweise als *Page Administrator*) oder ohne entsprechendes *Access Token* nicht zugegriffen werden. Insgesamt bleiben 32 *Fields*, die allgemein anwendbar und ohne zusätzliche Berechtigung verfügbar sind. [Fac17e]

Die allgemein verfügbaren *Fields* eines *Page Node* sind in Tabelle 2.1 dargestellt. Eine Beschreibung aller *Fields* eines *Page Node* findet sich in Tabelle A.7 im Anhang.

Tabelle 2.1: Allgemein verfügbare *Fields* eines *Page Node* (nach [Fac17e])

Field	Type	Description <sup>2</sup>
about	string	Informationen über die Seite
best_page	Page	Diejenige Seite auf FACEBOOK, welche dem Konzept, das diese Seite repräsentiert, am besten entspricht
can_checkin	bool	Angabe, ob die <i>Checkin</i> -Funktion dieser Seite aktiviert ist
can_post	bool	Angabe, ob der Nutzer der aktuellen Sitzung auf der Seite <i>posten</i> kann
category	string	Kategorie der Seite
category_list	list<PageCategory>	Unterkategorien der Seite
checkins	unsigned int32	Anzahl der <i>Checkins</i> an diesem Ort
contact_address	MailingAddress	Post- oder Besucheranschrift
context	OpenGraphContext	Sozialer Kontext dieser Seite
cover	CoverPhoto	Titelbild der Seite
current_location	string	Aktueller Standort der Seite
description	string	Beschreibung der Seite
display_subtext	string	Angezeigter Subtext der Seite
displayed_message_response_time	string	Geschätzte Antwortzeit der Seite
emails	list<string>	E-Mail-Adressen aus dem <i>Field about</i>
fan_count	unsigned int32	Anzahl der Nutzer, welche die Seite <i>ge-liked</i> haben
featured_video	Video	Besonders platziertes Video der Seite
general_info	string	Allgemeine Angaben der Seite
id	numeric string	<i>Page ID</i> der Seite
impressum	string	Informationen über den Herausgeber
is_community_page	boolean	Angabe, ob die Seite durch eine <i>Community</i> betreut wird
link	string	<i>URL</i> der Seite auf FACEBOOK
name	string	Titel der Seite
overall_star_rating	float	Gesamtbewertung der Seite
phone	string	Telefonnummer der Seite
rating_count	unsigned int32	Anzahl der Bewertungen
single_line_address	string	Adresse der Seite ohne Umbrüche
start_info	PageStartInfo	Informationen über den Start der Seite
talking_about_count	unsigned int32	Anzahl der Nutzer, welche innerhalb der vergangenen sieben Tage mit der Seite interagiert haben
username	string	Alias für FACEBOOK- <i>URL</i> der Seite
verification_status	string	Informationen über die Authentizität der Seite
website	string	<i>URL</i> der <i>Website</i> der Seite

<sup>2</sup> Für ein besseres Verständnis wurden die *Descriptions* hier übersetzt.

**Edges** Ein FACEBOOK *Page Node* kann potenziell über 60 verschiedene Typen von *Edges* jeweils mit einem oder mehreren anderen *Nodes* verbunden sein. Einige der *Edges* sind – analog zu den *Fields* – nur für bestimmte Arten von Seiten (beispielsweise Shopping-Angebote) oder spezielle Anwendungsfälle (beispielsweise eine mobile *App*) anwendbar. Auf einige *Edges* kann – ebenso wie bei den *Fields* – ohne zusätzliche Berechtigung (beispielsweise als *Page Administrator*) oder ohne entsprechendes *Access Token* nicht zugegriffen werden. Insgesamt gibt es neun<sup>3</sup> *Edges*, die allgemein anwendbar und ohne zusätzliche Berechtigung verfügbar sind. [Fac17e]

Die allgemein verfügbaren *Edges* eines *Page Node* sind in Tabelle 2.2 dargestellt. Eine Aufstellung aller *Edges* eines *Page Node* findet sich in Tabelle A.8 im Anhang.

Tabelle 2.2: Allgemein verfügbare *Edges* eines *Page Node* (nach [Fac17e])

Edge	Type	Description <sup>4</sup>
albums	list<Album>	Fotoalben dieser Seite
events	list<Event>	Veranstaltungen dieser Seite
likes	list<Page>	Seiten, welche diese Seite <i>geliked</i> haben
photos	list<Photo>	Fotos dieser Seite
picture	Picture	Profilbild dieser Seite
posts	Post[]	Alle Posts (inklusive Statusaktualisierungen), die von oder auf dieser Seite veröffentlicht wurden
tagged	list<Post>	Alle öffentlichen Fotos, Videos und Posts, in denen diese Seite markiert wurde
video_lists	list<VideoList>	Wiedergabelisten für Videos dieser Seite
videos	list<Video>	Videos dieser Seite

Über jede seiner *Edges* kann ein *Page Node* potenziell mit einem oder mehreren weiteren *Nodes* verbunden sein, deren Typ jeweils durch den Typ der betreffenden *Edge* bestimmt wird. Die mit einem *Page Node* verbundenen *Nodes* wiederum besitzen eigene spezifische *Edges*, über welche sie mit weiteren *Nodes* verbunden sein können. Die daraus potenziell entstehende Netzwerkstruktur – mit einer maximalen Pfadlänge von 1 (*1-Schritt-Umgebung*) bzw. 2 (*2-Schritt-Umgebung*) von dem *Page Node* ausgehend – ist in den Abbildungen 2.1 und 2.2 dargestellt [THS11, S. 101ff.].

<sup>3</sup> Formal gibt es zehn dieser *Edges*, allerdings liefern die *Edges feed* und *posts* nach Aussage der *Graph-API*-Dokumentation und eigenen Untersuchungen identische Ergebnisse. Wegen des eindeutigeren Begriffs wird in dieser Arbeit *posts* verwendet.

<sup>4</sup> Für ein besseres Verständnis wurden die *Descriptions* hier übersetzt.

Bei der Interpretation der Netzwerkstruktur ist zu beachten, dass die abgebildeten Knoten (*Nodes*) jeweils stellvertretend für eine potenziell sehr große Anzahl von Objekten stehen, die mit dem jeweiligen Typ einer Kante (*Edge*) verbunden sein können. Typischerweise sind auf einer *Page* viele *Posts* und *Photos* veröffentlicht und eine *Page* wurde von mehreren anderen *Pages* *geliked*. Die Knoten und Kanten in den Abbildungen sind daher als Typklassen zu interpretieren, von denen jeweils viele Instanzen existieren können.

Bereits diese vereinfachte Darstellung zeigt eindrucksvoll die enorme Komplexität und Dimension des Ausschnitts aus dem sozialen Netzwerk, die schon bei geringen betrachteten Distanzen innerhalb des FACEBOOK *Social Graph* sichtbar werden.

Die *Fields* und *Edges* der mit dem *Page Node* verbundenen *Nodes* werden hier nicht explizit betrachtet. Die Eigenschaften der für die Entwicklung des Goldstandards relevanten Knotentypen sind im Anhang A zusammengestellt und werden später bei der Auswahl von Kandidaten möglicher *Features* für die *Klassifizierung* von FACEBOOK *Pages* berücksichtigt (vgl. Abschnitt 4.1).

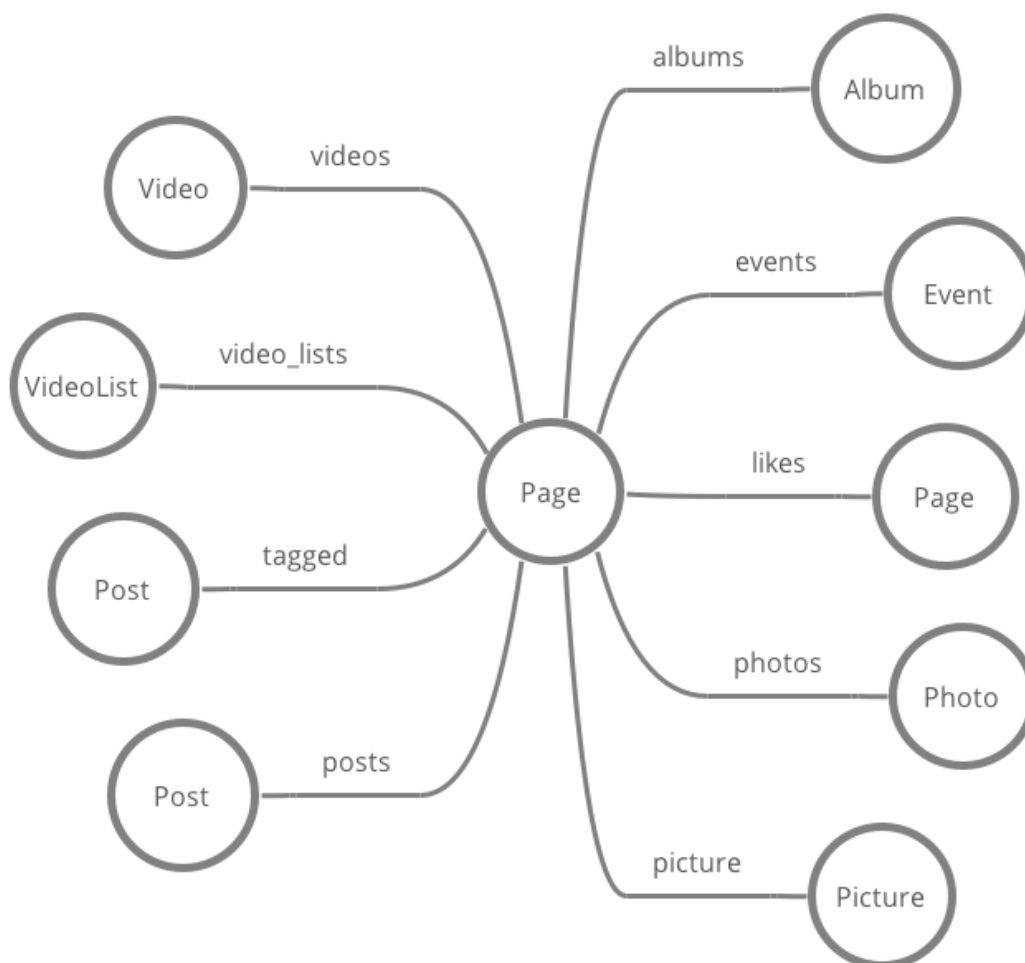


Abbildung 2.1: *Page Node* mit *Edges* und potenziell verbundenen *Nodes* (1-Schritt-Umgebung)



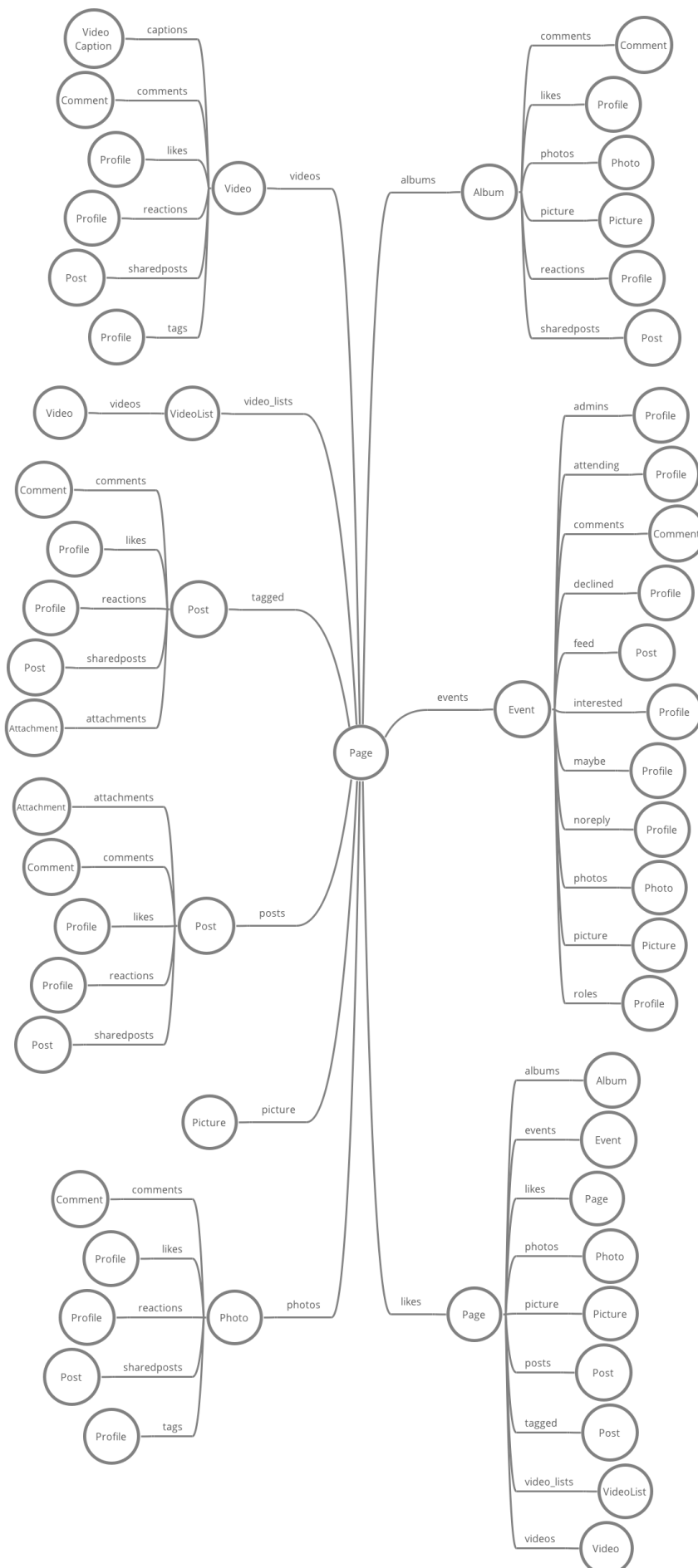


Abbildung 2.2: *Page Node* mit *Edges* und potenziell verbundenen *Nodes* (2-Schritt-Umgebung)



## 3 Detektion von Gefährdern

Die Detektion von Gefährdern im Kontext dieser Arbeit lässt sich als Problem der automatischen oder maschinellen *Klassifizierung* von Akteuren im sozialen Online-Netzwerk beschreiben. Klassifizierung bedeutet hier die Einordnung der Profile von persönlichen oder institutionellen Akteuren in die vorgegebenen Klassen *Gefährder* und *Nicht-Gefährder*, ein sogenanntes binäres Klassifizierungsproblem (siehe Abschnitt 3.2).

Der Begriff des *Gefährders* ist ein Arbeitsbegriff der Sicherheitsbehörden von Bund und Ländern in Deutschland für den Phänomenbereich der politisch motivierten Kriminalität. Ein äquivalenter Begriff, der von Polizeibehörden im Ausland verwendet wird, ist nicht bekannt. [Deu17a] Der Begriff ist bisher nicht legal definiert – es existiert lediglich eine Definition der ARBEITSGEMEINSCHAFT DER LEITER DER LANDESKRIMINALÄMTER UND DES BUNDESKRIMINALAMTS (AG KRIPO) aus dem Jahr 2004 [Wis17]:

*„Ein Gefährder ist eine Person, bei der bestimmte Tatsachen die Annahme rechtfertigen, dass sie politisch motivierte Straftaten von erheblicher Bedeutung, insbesondere solche im Sinne des § 100a der Strafprozessordnung (StPO)<sup>5</sup>, begehen wird.“*

Diese Definition ist jedoch rechtlich nicht verbindlich und der Begriff *Gefährder* wird von verschiedenen Sicherheitsbehörden in Bund und Ländern teilweise unterschiedlich definiert und verwendet. Letztendlich soll die Figur des Gefährders der Früherkennung *möglicher* Straftäter dienen, also von Personen, die zukünftig Straftaten begehen *können*. [Wis08, Wis17]

### 3.1 Merkmale von Gefährdern

Im Kontext dieser Arbeit geht es um die Detektion von Gefährdern in sozialen Online-Netzwerken. Um Gefährder beobachten und bei verdächtigen Aktivitäten genauer untersuchen zu können, müssen diese automatisiert erkannt und von unverdächtigen Profilen differenziert – das heißt als Gefährder klassifiziert – werden. Profile und Strukturen in sozialen Online-Netzwerken spiegeln – zu einem gewissen Teil – ein Abbild der realen (physischen) Welt wider (vgl. Abschnitt 2.3.2). Um geeignete *Features* für die Klassifizierung von Gefährdern in sozialen Online-Netzwerken zu finden, können daher zunächst deren Merkmale und Eigenschaften in der realen Welt herangezogen werden. [DACP15]

<sup>5</sup> Im hier referenzierten § 100a der StPO ist unter anderem ein Katalog schwerer Straftaten definiert, bei denen die Überwachung und Aufzeichnung der Telekommunikation (TKÜ) zulässig ist. [StP17a]

Problematisch bei der Analyse der Charakteristik von Gefährdern ist – neben der ohnehin schwierigen Definition des Begriffes – die schlechte und sich teilweise widersprechende Informationslage. Grundsätzlich ist die Beurteilung von Gefährdern Ländersache und „alle derzeit erfassten 'Gefährder' und 'relevanten Personen' wurden von den Polizeien der Länder in eigener Zuständigkeit eingestuft.“ [Deu16] Die Beurteilung einer Person in Hinblick auf eine Einstufung als Gefährder erfolgt jeweils in einem individuellen Prozess je nach Zuständigkeit durch verschiedene Polizeibehörden der Länder in Zusammenarbeit mit den Landesämtern für Verfassungsschutz. Informationen darüber, welche Tatsachen und Kriterien unter welchen Umständen zu einer Einstufung als Gefährder führen, sind klassifiziert und daher öffentlich nicht zugänglich. Voraussetzung für eine Einstufung als Gefährder sind jedoch weder eine konkrete Straftat, noch Anhaltspunkte für deren Vorbereitung oder Versuch, noch Hinweise auf eine konkrete Gefahr für die öffentliche Sicherheit.<sup>6</sup> [Wis08, Deu17a, Deu17b, Kre17]

Aus Interviews, die KRETSCHMANN in den Jahren 2011 und 2012 mit Beamten des Bundeskriminalamtes und verschiedener Landeskriminalämter zum Thema *Gefährder* führen konnte, ergeben sich zumindest einige Hinweise auf das Vorgehen und mögliche Kriterien für die Einstufung von Gefährdern. Um das Gefahrenpotenzial bestimmter Personen festzustellen, werden *Gefahrenermittlungen* durchgeführt und ein sogenanntes *Personagramm* erstellt. Es werden „[...] Familienstand, Staatsangehörigkeit, aktuelle und frühere Wohnsitze, Konten und Kontobewegungen, Kraftfahrzeugbesitze, persönliche Telefonnummern und das persönliche Umfeld erfasst [...], gegebenenfalls auch der Asylstatus und der Weg der Einreise nach Deutschland.“ [Kre17] Die Erkenntnisse der Sicherheitsbehörden lassen sich grob drei Lebensbereichen zuordnen (vgl. [Kre17, Lan17b, Bun17]):

- **Biografische Angaben** – unter anderem Herkunft, Ausbildung, Werdegang, besondere Kenntnisse (Sprachen, Waffen, Sprengstoff),
- **Persönliche Situation** – unter anderem Beziehungsstatus, Wohnung, Finanzen, Arbeitsplatz,
- **Umfeld und Verhalten** – unter anderem Kontakte zu Familie, Personen und Gruppierungen einer bestimmten Szene, auffälliges Verhalten.

Die beschriebenen Aspekte geben zwar keinen Aufschluss über die konkreten – für eine Einstufung als Gefährder tatsächlich angewendeten – Kriterien und deren relevante Ausprägungen, liefern aber zumindest Hinweise auf Typ und Herkunft von Merkmalen und Eigenschaften, die für eine Einstufung herangezogen werden.

Die genannten Lebensbereiche und deren relevante Merkmale finden sich wenigstens teilweise in den Attributen sowie der Position und Verknüpfung der betreffenden Person

<sup>6</sup> Auf die weitergehende Betrachtung der juristisch problematischen Konstellation möglicher Grundrechtseingriffe wegen antizipiertem strafbarem Verhalten bei Personen, die als Gefährder eingestuft wurden, ohne „zureichende tatsächliche Anhaltspunkte“ für das Vorliegen einer verfolgbaren Straftat (§ 152 II StPO, [StP17b]), wird hier verzichtet. (vgl. dazu [Bö11, Kre17])

in sozialen Online-Netzwerken wieder. Wie in Abschnitt 2.1 ausgeführt, sind die Attribute eines Akteurs mit seiner Position und seiner Verknüpfung sowie mit seinem Verhalten im sozialen Netzwerk korreliert. Ohne die Kausalität der Korrelationen im Einzelfall zu betrachten, können sich so zusätzlich zu den expliziten Attributen und Verbindungen einer Person oder Organisation aus dem beobachteten Verhalten eines Akteurs im sozialen Online-Netzwerk (Teilen, Bewerten, Kommentieren usw.) weitere Informationen beispielsweise über gemeinsame Einstellungen, Ansichten, Stimmungen oder Meinungen ableiten lassen (siehe Abschnitt 3.3).

Unter dieser Annahme könnten beispielsweise bekannte FACEBOOK-Profile, deren Urheber als Gefährder eingestuft oder entsprechend beobachtet werden, für das Training von Klassifikatoren genutzt werden, die Aktivitäten von Gefährdern in sozialen Online-Netzwerken identifizieren sollen. Um einen solchen Klassifikator für ein bestimmtes Bedrohungsszenario zu entwickeln, müssen relevante domänenspezifische Profile und *Features* als Trainings- und Testdaten ausgewählt sowie darauf basierend ein Klassifizierungsalgorithmus trainiert werden. [SHGL15] Im folgenden Abschnitt 3.2 werden diese Aspekte näher betrachtet.

## 3.2 Exkurs: Klassifizierung und Maschinelles Lernen

Die *automatische Klassifizierung* ist ein Teilgebiet des *Maschinellen Lernens* (*Machine Learning*). Beim *Maschinellen Lernen* werden durch ein künstliches System anhand von Beispielen Muster und Gesetzmäßigkeiten erkannt, das heißt durch einen Algorithmus „gelernt“. Ziel des *Maschinellen Lernens* ist im Allgemeinen die *Generalisierung* eines (verborgenen) Konzeptes, so dass das System nach einer Trainingsphase anhand der gelernten Zusammenhänge auch unbekannte Daten beurteilen kann. *Maschinelles Lernen* teilt viele Methoden und Algorithmen mit dem *Data Mining*, wobei beim *Data Mining* primär *neue* Muster in Daten gefunden werden sollen, während beim *Maschinellen Lernen* der Fokus darauf liegt, *bekannte* Muster in neuen Daten wiederzuerkennen. [Agg14, S. 2] [Rei16]

Klassifizierung ist eine Form des *überwachten Lernens* (*Supervised Learning*). Im Gegensatz zum *unüberwachten Lernen* (*Unsupervised Learning*), wie beispielsweise das *Clustering* von Daten, sind beim überwachten Lernen nicht nur die Eingabedaten, sondern auch die zugehörigen Ausgabewerte bekannt und vorgegeben. In einer Trainingsphase wird der *Klassifizierungsalgorithmus* (*Lernmodell*, *Klassifikator*) auf die – im Allgemeinen verborgenen – Zusammenhänge zwischen Eingabedaten und Ausgabewert trainiert. Die Eingabedaten repräsentieren die Merkmale (*Features*) der zu klassifizierenden Objekte und die Ausgabewerte entsprechen den verschiedenen Zielklassen (*Labels*), welchen die Objekte zugeordnet werden sollen (siehe Abbildung 3.1). [Rei16]

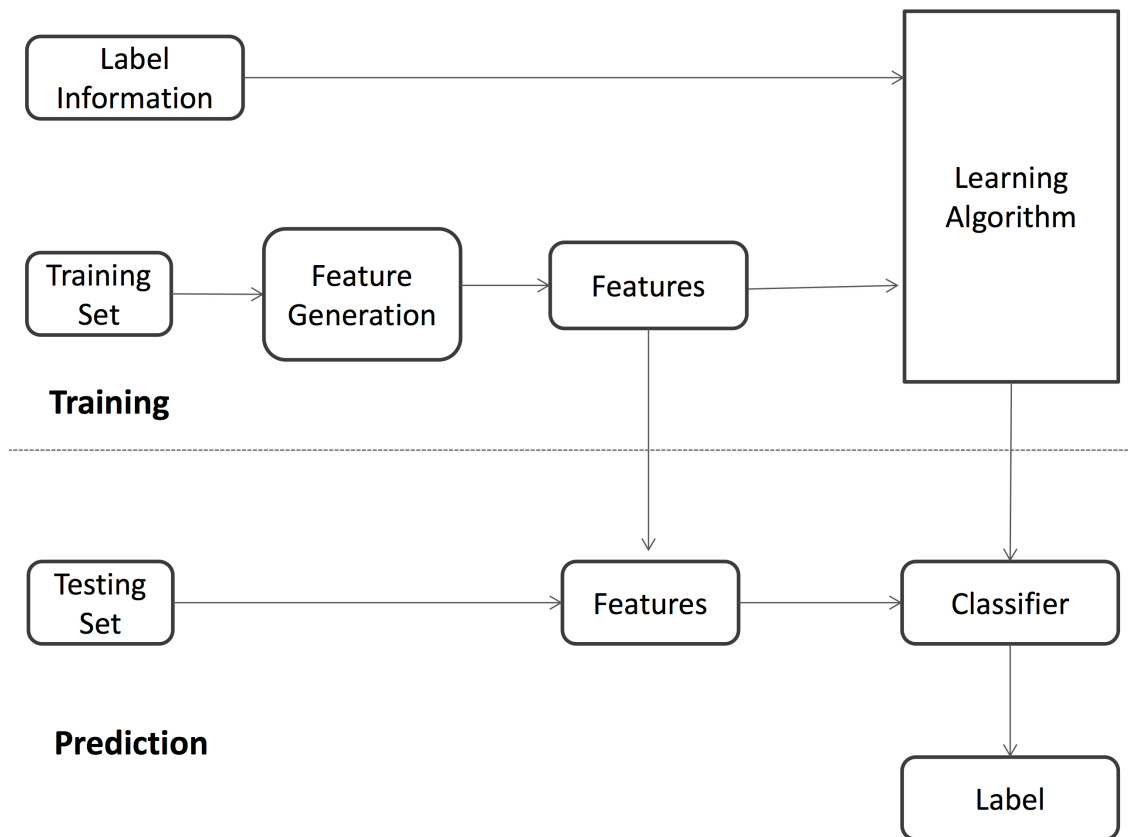


Abbildung 3.1: Allgemeiner Prozess der Klassifizierung (nach [Agg14, S. 40])

### 3.2.1 Finden geeigneter Features

*Features* können allgemein als gemeinsame Eigenschaften oder Merkmale der zu klassifizierenden Objekte beschrieben werden, die geeignet sein können, die verschiedenen Objekte hinsichtlich der Zielklassen ausreichend zu differenzieren. *Features* können einzelne Parameter, aber auch komplexe Strukturen der Eingabedaten sein. Eine häufige Herausforderung für die Klassifizierung ist die hohe Dimensionalität der Merkmale der ursprünglichen Eingabedaten. Dies verursacht einerseits sehr große Datenmengen und hohe Anforderungen an die Berechnung der Klassifizierungsalgorithmen (Datendurchsatz, Rechenleistung, Speicherplatz) und kann andererseits die Überanpassung (*Overfitting*) eines Lernmodells begünstigen. Als *Overfitting* wird das Phänomen bezeichnet, bei dem ein Klassifikator stärker die individuellen Ausprägungen der Trainingsdaten lernt, statt die zugrundeliegenden differenzierenden Charakteristika der Zielklassen. Dies führt zu sehr guten Klassifizierungsergebnissen in der Trainingsphase, reduziert aber die Generalisierbarkeit des Klassifikators und führt damit zu einer geringeren Genauigkeit bei der Klassifizierung neuer, unbekannter Eingabedaten. [IKT05] (siehe dazu auch Abschnitt 3.2.4)

Um die Dimensionalität der Daten zu reduzieren und geeignete *Features* für die Klassifizierung zu finden, werden häufig Methoden der *Feature Selection* und *Feature Extraction* angewendet [For03, IKT05]:

**Feature Selection** Bei der *Feature Selection* werden Merkmale des Datensatzes entfernt, die für die Klassifizierung als nicht relevant eingeschätzt werden. Es existieren verschiedene statistische Verfahren und Metriken, mit deren Hilfe für jedes einzelne *Feature* unabhängig voneinander ein *Score* (Punktwert) ermittelt kann, anhand dessen eine Rangfolge der *Features* entsprechend ihrem Anteil an der korrekten Klassifizierung (Stärke der Korrelation zwischen Eingabe- und Ausgabewert) gebildet wird. Die am höchsten bewerteten *Features* werden ausgewählt und für den weiteren Prozess verwendet (siehe Abbildung 3.2). [Agg14, S. 40ff.]

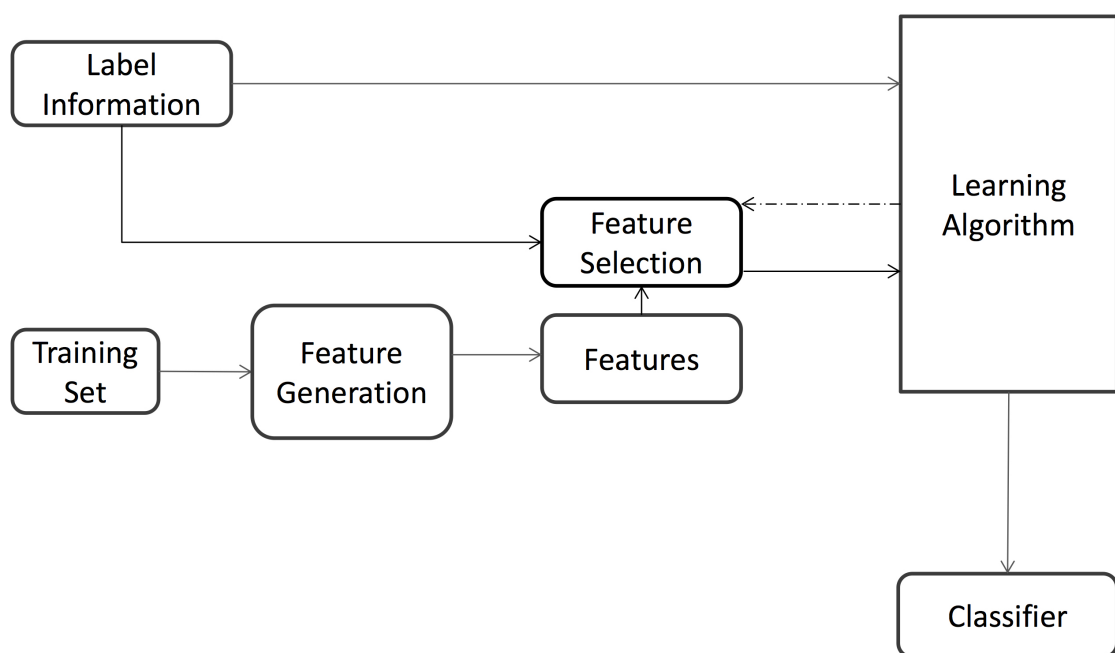


Abbildung 3.2: Feature Selection im Klassifizierungsprozess (nach [Agg14, S. 42])

**Feature Extraction** Bei der *Feature Extraction* (auch *Feature Transformation*) werden aus den vorhandenen (häufig vorselektierten) *Features* neue *Features* mit besser diskriminierenden Eigenschaften – meist einhergehend mit gleichzeitiger Reduktion der Dimensionalität der Daten – gebildet (siehe Abbildung 3.3). Ein häufiger Ansatz für die *Feature Extraction* ist die *Hauptkomponentenanalyse* (*Principal Component Analysis – PCA*). Die *Hauptkomponentenanalyse* versucht, einen hochdimensionalen Merkmalsraum durch *Hauptachsentransformation* in einen Merkmalsraum mit kleinerer Dimension zu projizieren. Dabei wird die in den Ausgangsvariablen verborgene *Redundanz* eliminiert und die diskriminierende Information durch eine geringere Anzahl aussagekräftiger *Features* (*Hauptkomponenten*) abgebildet, so dass diese einen größtmöglichen Anteil der Varianz der Ausgangsdaten erklären. [Agg14, S. 38]

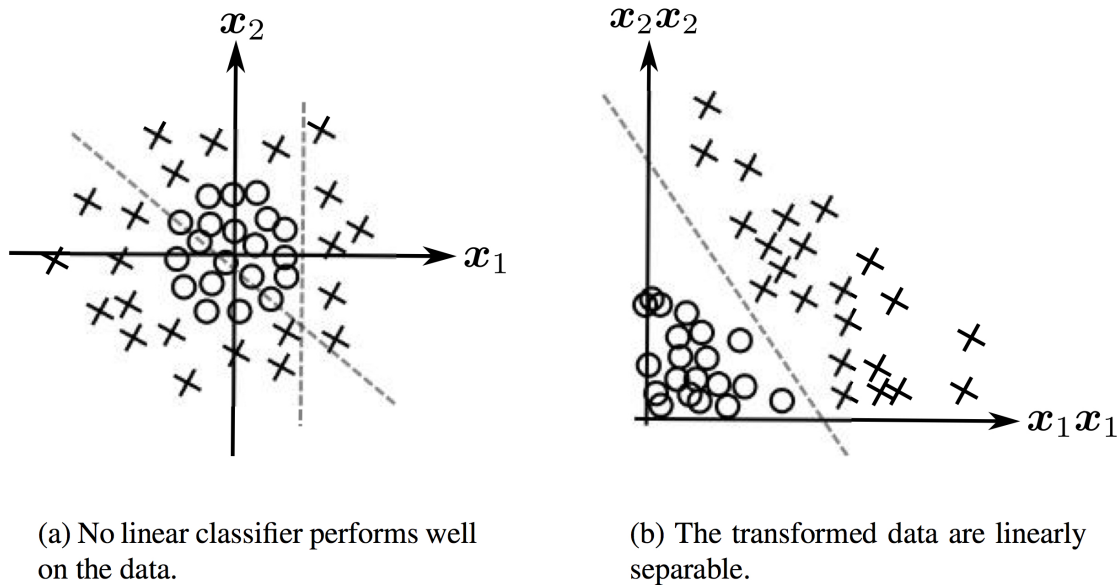


Abbildung 3.3: Feature Transformation für bessere lineare Trennbarkeit (nach [Agg14, S. 194])

### 3.2.2 Wahl des Klassifizierungsverfahrens

Neben der Vorverarbeitung der Daten und dem Finden geeigneter *Features* sind die Wahl des passenden Klassifizierungsverfahrens und eines effizienten Klassifikators entscheidend für die Leistungsfähigkeit (*Performance*) der Klassifizierung. Übliche und häufig genutzte Ansätze für die Klassifizierung sind [LLS00] [Agg14, S. 6ff.]:

- **Entscheidungsbäume** – streng hierarchische Partitionierung der Daten durch mehrstufiges Aufteilen der Daten anhand eines oder mehrerer Attribute,
- **Regelbasierte Verfahren** – ähnlich Entscheidungsbäumen, aber weniger strikt – Überlappungen beim Aufteilen der Daten sind zulässig,
- **Statistische Verfahren** – Klassifizierung auf Basis von Wahrscheinlichkeiten,
- **Instanzbasiertes Lernen** – Instanzen der zu klassifizierenden Testdaten trainieren gleichzeitig das Modell (*Lazy Learning*),
- **Support Vector Machine** – Trennen zweier Klassen durch eine lineare Hyperebene im Attributraum mit größtmöglichem Abstand zwischen den Klassengrenzen,
- **Neuronale Netzwerke** – Simulation biologischer Prozesse des menschlichen Gehirns durch ein Netzwerk künstlicher Neuronen und Synapsen.

Darüber hinaus existieren weitere – auch hybride – Verfahren und die Weiterentwicklung und Optimierung von Klassifizierungsalgorithmen ist ein intensiv bearbeitetes Forschungsgebiet [Agg14, S. 4]. Die Auswahl einer geeigneten Klassifizierungsmethode und eines effizienten Klassifikators verlangt einige Erfahrung und ist – vor allem bei neuartigen Aufgabenstellungen – nicht selten in den experimentellen Gesamtprozess der Klassifizierung eingebunden. Die Wahl des passenden Klassifizierungsverfahrens



hängt unmittelbar von der konkreten Aufgabenstellung, aber auch von verschiedenen weiteren Rahmenbedingungen ab; die wichtigsten Aspekte sind [LLS00, Kot07] [Agg14, S. 16ff.]:

- **Fachdomäne und Art des Klassifizierungsproblems** – beispielsweise Biologie, Medizin, Marketing, Soziale Netzwerke, Information Retrieval,
- **Qualität der Daten** – beispielsweise fehlende Werte, Redundanz, Rauschen,
- **Natur der Daten** – beispielsweise Texte, Dokumente, Multimedia, Zeitreihen, Graphen, Datenströme,
- **Typen und Skalenniveau der Attribute** – nominal, ordinal, kardinal,
- **Anforderungen an Geschwindigkeit** – sowohl beim Training als auch bei der Klassifizierung,
- **Ressourcenbedarf** – beim Training und während der Laufzeit des Klassifikators (Rechenzeit, Speicherbedarf),
- **Anforderungen an Skalierbarkeit** – großer Umfang der Daten oder Attribute.

Um die *Performance* für die Auswahl eines Klassifikators mit anderen Verfahren bzw. Konfigurationen zu vergleichen, muss dieser evaluiert werden. Bei der *Evaluation* werden bestimmte Maßzahlen bestimmt, um die Güte eines Klassifikators zu beurteilen (siehe Abschnitt 3.2.4).

### 3.2.3 Kombinieren von Klassifikatoren

Die *Performance* von Klassifikatoren hängt stark davon ab, wie gut ein *Lernmodell* auf die konkrete Aufgabenstellung passt und trainiert ist. Es liegt in der Natur von *Social Media*, dass die Inhalte eine sehr große thematische und mediale Bandbreite haben (siehe Abschnitte 2.3.1 und 3.3). Um trotzdem zu robusten Ergebnissen zu gelangen, hat es sich bewährt, mehrere Klassifikatoren, die jeweils auf einen speziellen Anwendungsfall trainiert sind, zu kombinieren und über einen *Voting*-Mechanismus eine gemeinsame Entscheidung für die Klassifizierung abzuleiten. Bei diesem sogenannten *Ensemble Learning* ist die größte Herausforderung, eine geeignete Kombination von Klassifikatoren und ein passendes *Voting*-Verfahren zu entwickeln (*Classifier Committee Construction*). [Agg14, S. 484ff.]

Einen auch im Kontext dieser Arbeit interessanten Vorschlag haben SPRANGER et al. auf Basis eines künstlichen Immunsystems entwickelt, mit dem strafrechtlich relevante Informationen in sozialen Online-Netzwerken gesammelt werden können. Ein *Classifier Ensemble* übernimmt hier die Rolle, Kandidatenprofile in sozialen Online-Netzwerken zu identifizieren, die dann durch sogenannte *Analysten* des künstlichen Immunsystems beobachtet werden (siehe Abbildung 3.4). Um die Klassifikatoren des Ensembles entsprechend zu trainieren, sollen Profile bekannter Gefährder oder Straftäter aus sozialen Online-Netzwerken herangezogen werden. [SHGL15, SBHL17]

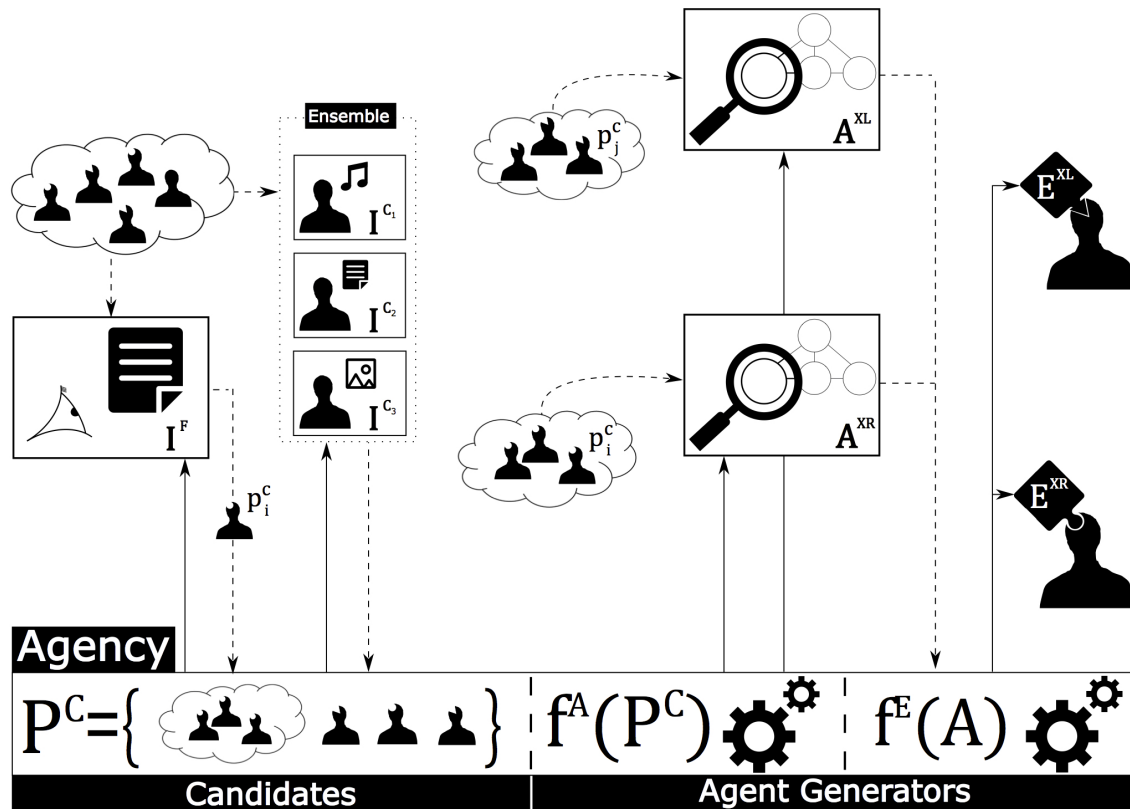


Abbildung 3.4: Schematische Struktur eines künstlichen Immunsystems, bei dem Kandidatenprofile  $P^C$  durch ein Ensemble sogenannter Informanten  $I^C$  identifiziert und anschließend durch Analysten  $A$  beobachtet werden. [SBHL17]

Für die Ermittlung des gemeinsamen Klassifizierungsergebnisses (*Rank*, *Score* oder *Label*) gibt es je nach Art der Klassifikatoren und Anwendungsfälle verschiedene Ansätze, wie beispielsweise eine einfache Mehrheitsentscheidung oder – optional gewichtete – Linearkombinationen von *Score*- oder *Ranking*-Werten. [AZ12, S. 209f.] [Agg14, S. 23ff.] ZHANG et al. und KITTLER et al. beschreiben ausführlich die Entwicklung und Konstruktion von *Ensemble Systems* und gehen auch detailliert auf die Anforderung ein, heterogene Daten aus verschiedenen Quellen mit unterschiedlichen Formaten, Dimensionen und Strukturen zu verarbeiten (sogenannte *Data Fusion*). Bei komplexen Anforderungen können die einzelnen Klassifikatoren eines Ensembles selbst ebenfalls wiederum durch ein Ensemble gebildet werden. [ZM12, S. 1ff., S. 21f.] [KR00, S. 362ff.]

### 3.2.4 Evaluation von Klassifizierungsalgorithmen

Für die Lösung eines gegebenen Klassifizierungsproblems existieren oft verschiedene Verfahren und Parameter. Um festzustellen, wie gut sich das Problem überhaupt lernen lässt und um die am besten geeignete Konfiguration für ein Lernmodell zu finden, muss dessen Performance bestimmt und mit anderen Konfigurationen verglichen werden. Darüber hinaus dient die Evaluation eines Klassifikators nicht nur dem Fin-

den optimaler Verfahren und Parameter, sondern auch der Feststellung seiner Güte. So können Aussagen darüber getroffen werden, in welcher Höhe Fehler eines bestimmten Typs bei einem Klassifikator zu erwarten sind, um dessen Ergebnisse im praktischen Einsatz überhaupt sinnvoll interpretieren und entsprechende Schlüsse ziehen zu können. [Agg14, S. 30f.]

**Evaluierungsstrategie** Für das Training eines Klassifikators (Lernen eines Modells) werden typischerweise Trainingsdaten benötigt, um beispielsweise die Knoten und Schwellwerte für Entscheidungsbäume oder Stützvektoren einer *Support Vector Machine* zu ermitteln. Im einfachsten Fall werden deshalb die verfügbaren gelabelten Daten durch zufällige Auswahl von Instanzen in zwei Teile geteilt, so dass ein Teil für das Training (Trainingsdaten) und der andere Teil für die Evaluation (Testdaten) genutzt werden (**Hold-out Evaluation**). Häufig werden zwei Drittel der Daten für das Training und ein Drittel für den Test verwendet. Nachteil dieses Evaluierungsdesigns kann sein, dass unter Umständen essenzielle Instanzen für das Training verloren gehen oder die Testergebnisse wegen ungünstig gewählter Testinstanzen zu hoch bzw. zu niedrig ausfallen. Deshalb wird die *Hold-out Evaluation* vor allem bei großen Datenmengen durchgeführt oder der Mittelwert mehrerer Durchläufe mit jeweils neu gewählten Trainings- und Testdaten verwendet. Bei der häufig genutzten **k-fold Cross Validation** werden die Daten in  $k$  gleich große Teile gesplittet, typischerweise liegt  $k$  zwischen 5 und 10. Anschließend wird der Algorithmus  $k$ -mal trainiert und evaluiert, wobei jeweils  $k - 1$  Teile für das Training und 1 Teil für die Evaluation verwendet werden. Vorteil dieser Strategie ist, dass jede Instanz des Datensatzes und damit jede Ausprägung der *Features* genau einmal evaluiert wird. [Agg14, S. 634f.]

**Performancemaße** Um einen Klassifikator zu evaluieren, werden üblicherweise Kennzahlen bestimmt, die seine **Effektivität** messen, das heißt die Fähigkeit, richtig zu klassifizieren. Häufig gemessene Maßzahlen, welche die Effektivität eines Modells beschreiben, sind *Precision*, *Recall*, *Accuracy* und *Error*. Die **Precision** (*Genauigkeit*) gibt an, wie viele der einer bestimmten Klasse zugeordneten Objekte dieser Klasse korrekt zugeordnet wurden. Der **Recall** (*Sensitivität*) misst den Anteil der einer bestimmten Klasse korrekt zugeordneten Objekte an der Gesamtheit der tatsächlich dieser Klasse zugehörigen Instanzen. Die **Accuracy** (*Korrektklassifikationsrate*) misst den Anteil aller korrekt klassifizierten Objekte an der Menge aller klassifizierten Objekte. Der **Error** (*Falschklassifikationsrate*) hingegen gibt an, wie groß der Anteil aller falsch klassifizierten Objekte an der Menge aller klassifizierten Objekte ist. Die beschriebenen Maßzahlen beeinflussen sich gegenseitig und sind isoliert nicht für eine fundierte Beurteilung der Güte eines Klassifikators geeignet. Besonders wenn die Zielklassen nicht gleichmäßig über die Menge aller Objekte verteilt sind und der Anteil einer Klasse besonders hoch oder gering ist, sind einzelne Performancemaße nicht sehr aussagekräftig. Um die Güte eines Klassifikators besser durch eine einzelne Kennzahl beschreiben zu können, wurden kombinierte Maßzahlen entwickelt, von denen das *F-Maß* die häufigste Verwendung findet. Das **F-Maß** wird durch das gewichtete harmonische Mittel von *Precision* und

*Recall* bestimmt, wobei durch einen zusätzlichen Parameter das Gewicht der beiden Kennzahlen im kombinierten Maß je nach Anforderungen und Anwendungsfall beeinflusst werden kann. [Seb01, IKT05] [Agg14, S. 636ff.]

**Weitere Aspekte** In bestimmten Anwendungsfällen – beispielsweise wenn die gemessene Effektivität verschiedener Klassifikatoren sehr nah beieinander liegt – kann es sinnvoll sein, darüber hinaus alternative Aspekte eines Lernmodells zu betrachten. Unter dem Gesichtspunkt der **Effizienz** eines Klassifikators können unter anderem die durchschnittliche Zeit für das Lernen einer Klasse auf Basis eines Trainingsdatensatzes (*Training Efficiency*) oder die durchschnittliche Zeit für das Zuordnen eines neuen Objektes zu einer Klasse (*Classification Efficiency*) bestimmt werden. Eine weitere Alternative zur Effektivität ist der **Nutzwert** (*Utility*), bei dem die ökonomischen Konsequenzen bestimmter Klassifizierungsentscheidungen (monetär) bewertet und der korrespondierende Gewinn oder Verlust zusätzlich zur Evaluation eines Klassifikators herangezogen werden. [Seb01, Kot07]

Sowohl die Auswahl geeigneter *Features* und Klassifizierungsverfahren, als auch das Training und die Evaluation der Lernmodelle (*Klassifikatoren*) erfordern einen *gelabelten Datenkorpus*, das heißt einen Satz von Eingabedaten für die Klassifizierung, für welche auch die zugehörigen Ausgabewerte (Zielklassen, *Labels*) bekannt sind. Ein solcher standardisierter Korpus gelabelter Echtdaten für eine bestimmte Domäne und einen definierten Anwendungsfall wird *Goldstandard* genannt. [SL17, S. 187] Die Entwicklung eines Goldstandards für die Detektion von Gefährdern in sozialen Netzwerken ist Gegenstand dieser Arbeit und wird in Kapitel 4 genauer betrachtet.

### 3.3 Herausforderungen bei der Klassifizierung im Kontext von Social Media

Mit der Vielzahl der verschiedenen Funktionsbereiche und Anwendungen im Umfeld von *Social Media* ergeben sich für die Klassifizierung von Daten eine Reihe besonderer Herausforderungen. Bei *Social Media* handelt es sich im wörtlichen Sinne um *User-generated Content*, also um Inhalte, die von jedem einzelnen der Milliarden von Internetnutzern weltweit erstellt, bearbeitet, geteilt, kommentiert und bewertet worden sein können. Aufgrund ihrer Herkunft und Genese lassen sich diese Inhalte naturgemäß weder akademisch korrekt und eindeutig ordnen, noch folgen sie allgemein gültigen Konventionen irgendeiner Art. Alles ist möglich und wird gemacht, was gedacht und technisch gelöst werden kann. Autoren, Themen und Inhalte sind sehr dynamisch, fluktuierend und volatil. (siehe zu den Merkmalen von *Social Media* auch Abschnitt 2.3.1) [MGN12]

Neben der inhaltlichen Bandbreite und Vielfalt von *Social Media* gibt es auch auf Grund des *multi-modalen* und *multi-medialen* Charakters der Kommunikation spezielle

Anforderungen, die bei der Klassifizierung von *Social-Media*-Inhalten relevant sind [BR14] [Agg14, S. 337ff.].

Allein die **technische Verarbeitung** des großen Volumens verschiedener und im Normalfall parallel verwendeter Medientypen wie Text-, Bild-, Video- und Audiodaten stellt eine erste Hürde dar. Limitierungen treten hier wegen der begrenzten Bandbreiten von Schnittstellen, Netzwerk und Rechner sowie hinsichtlich des benötigten Arbeits- und Datenspeichers auf. Die teilweise komplexen internen Strukturen und zeitveränderlichen Signalwerte der verschiedenen Medien- und Dokumenttypen bedeuten weitere Herausforderungen für die Vorverarbeitung (*Feature Selection*, *Feature Extraction*) und Klassifizierung der Multimedia-Daten. [Agg14, S. 337ff.]

Nicht nur die mediale, sondern auch die *thematische Vielfalt* von *Social Media* ist nahezu unbegrenzt. Das ist für die automatische Klassifizierung der Inhalte problematisch, da herkömmliche Lernmodelle für genau ein Fachgebiet, eine Sprache oder einen Typ von Text trainiert werden müssen, um gute Ergebnisse zu erzielen. Sobald diese **Klassifikatoren** dann **auf neue Gebiete angewendet** werden, verringert sich deren Performance drastisch (siehe dazu auch Abschnitt 3.2.3). [SL14a] [XZHC13] [MBR12]

Weiterhin liegt es in der Natur von *User-generated Content*, dass dieser in **Aspekten der Datenqualität** keinen genormten Standards oder Mindestanforderungen genügt. Im Allgemeinen muss bei der Verarbeitung der Daten aus *Social-Media*-Anwendungen davon ausgegangen werden, dass substantielle Informationen durch ein starkes Rauschen irrelevanter Daten verdeckt sein können. [MBR12] [PFML16, S. 171ff.]

Neben diesen technischen und methodischen Problemfeldern ergeben sich allgemeine Herausforderungen bei der maschinellen **Verarbeitung menschlicher Sprache** (*Natural Language Processing* – NLP). Nachdem Texte in ihre Bestandteile zerlegt und deren *grammatikalische Funktion* (Wortart) bestimmt wurden, müssen *syntaktische Strukturen*, *Semantik* (Bedeutung) und *Pragmatik* (Absicht, Motivation) erkannt und bewertet werden. *Ambiguitäten* (Mehrdeutigkeiten) auf Wort- und syntaktischer Ebene sowie *anaphorische Verbindungen* werden aufgelöst und *Präsuppositionen* (implizites Vorwissen) extrahiert. Besondere Schwierigkeiten bereiten auch das Erkennen von *Sarkasmus* und *Ironie* sowie die *Negation* von Begriffen und Aussagen. [MBR12] [ZM16, S. 39ff.]

Darüber hinaus gibt es speziell bei Texten aus *Social-Media*-Anwendungen und besonders sozialen Online-Netzwerken weitere **linguistische Besonderheiten**, welche die maschinelle Verarbeitung und Klassifizierung erschweren. Häufig haben diese Texte nur einen verhältnismäßig geringen Umfang – vor allem in sozialen Online-Netzwerken – und Wörter oder ganze Aussagen werden zusätzlich durch Abkürzungen und *Hashtags* sowie *Emojis* und *Emoticons* – oft auch mehrdeutig – kodiert und verkürzt. [MGN12, MBR12]

Gerade im forensisch relevanten Umfeld sind die verwendeten *Codes* sehr volatil und ohne ständig aktualisiertes Expertenwissen kaum zu entschlüsseln. Dieses Problem der sogenannten verborgenen Semantik (*Hidden Semantics*) kann durch domänen-spezifische *Ontologien* und *Topic Maps* beherrschbar gemacht werden. Das ist umso wichtiger, wenn der unmittelbare Kontext einer Nachricht wegen der vernetzten und verschachtelten Struktur der Konversationen in sozialen Netzwerken schlecht zu erkennen ist. [SSH<sup>+</sup>12, SHAL16] [SL17, S. 187]

Häufig zeichnen sich Beiträge in sozialen Online-Netzwerken – bei erkennbar differenzierten sprachlichen Fähigkeiten – durch das Ignorieren von Sprachkonventionen mit unorthodoxer Rechtschreibung, Grammatik und Zeichensetzung aus. Autoren in sozialen Online-Netzwerken verwenden darüber hinaus häufig Formulierungen und Schreibweisen aus Dialektsprachen oder gruppenspezifisches Szene-Vokabular (*Slang*). Gerade die jüngere Zielgruppe kommuniziert gern mehrsprachig, oft auch gemischt innerhalb eines Textes. [MBR12, AC12, SL14b]

Besonders bei sozialen Online-Netzwerken werden die beschriebenen Herausforderungen noch durch eine weitere ergänzt, deren Ursprung in der **Struktur sozialer Netzwerke** begründet ist: im Gegensatz zum linear-sequenziellen Charakter konventioneller Veröffentlichungen sind die Akteure – ebenso wie die publizierten Medien – in sozialen Online-Netzwerken typischerweise stark miteinander vernetzt. Darüber hinaus existieren verschiedenartige Verbindungen zwischen den Akteuren (unter anderem *Befreundet*, *Publiziert* *Markiert*, *Bewertet*, *Geteilt*, *Kommentiert*), die wiederum selbst verschiedenen Typs sein können (*User*, *Pages* usw.) (siehe Abschnitt 2.3). [Agg14, S. 399ff.]

## 4 Entwicklung des Goldstandards

Der Begriff *Goldstandard* stammt ursprünglich aus dem Finanzwesen und beschreibt ein Währungssystem, bei dem der Wert des umlaufenden Geldes direkt an den Wert und die Existenz physischer Goldreserven gekoppelt ist und die nationale Währung zu einem garantierten Wechselkurs in Gold getauscht werden kann. Der Goldstandard wurde in der zweiten Hälfte des 19. Jahrhunderts durch viele aufstrebende und wirtschaftlich starke Staaten eingeführt und als Grundlage eines extrem stabilen Finanzsystems betrachtet. [BR96] Der Begriff wurde später von der medizinischen Forschung aufgegriffen und beschreibt dort allgemein anerkannte Referenzmethoden oder -verfahren mit der höchsten verfügbaren Genauigkeit und Verlässlichkeit. [Med17] Im Bereich des *Maschinellen Lernens* wird mit dem Begriff Goldstandard ein von Experten manuell *annotierter* – mit *Labels* und weiteren Metainformationen versehener –, standardisierter Referenzdatensatz bezeichnet, der als einheitliche Basis für die Entwicklung, die Evaluation und den Vergleich von Lernverfahren und Lernmodellen dient. [WAMDP14] [SL17, S. 187]

Die Referenzdaten werden mit definierten *Labels* versehen und gegebenenfalls mit weiteren Metadaten annotiert, so dass diese für das Training und die Evaluation von Klassifikatoren genutzt werden können. Mit der Bestimmung von Performancemaßen können die Güte eines Klassifikators in Bezug auf den Referenzdatensatz bestimmt und verschiedene Klassifikatoren miteinander verglichen werden (siehe Abschnitt 3.2.4). Durch den Goldstandard werden kontrollierte Bedingungen definiert, welche die *Vergleichbarkeit*, *Nachvollziehbarkeit*, *Automatisierbarkeit*, *Wiederholbarkeit* und *Konsistenz* von Test und Evaluation verschiedener Verfahren sicherstellen.

### 4.1 Anforderungen

Die Daten des Goldstandards müssen so vorbereitet werden, dass sie durch Klassifizierungs- und Evaluationsverfahren maschinell verarbeitet werden können. Dazu gehören neben einem geeigneten Datenformat auch zusätzliche Metainformationen (beispielsweise Label für Zielklassen), mit denen die Daten manuell annotiert werden müssen. Die Annotation mit dem Wissen menschlicher Experten ist eine wichtige Voraussetzung für die effektive und effiziente Anwendung von Lernalgorithmen und muss so sorgfältig wie möglich und unter Beachtung der konkreten Aufgabenstellung erfolgen. [PS13, S. 2] Es hat sich gezeigt, dass bei einem gegebenen Klassifizierungsproblem die Güte des entwickelten Klassifikators hauptsächlich von der Qualität der gelabelten Trainingsdaten abhängt. Selbst wenn die gleiche Klassifizierungsmethode angewendet wird, kann die Performance der resultierenden Klassifikatoren erheblich voneinander abweichen, wenn diese mit unterschiedlichen Trainingsdaten gelernt wurden. [IKT05, Rei16]

Im Kontext dieser Arbeit umfasst der Goldstandard einen *gelabelten Echtdatenkorpus* repräsentativer, institutioneller Profile des sozialen Online-Netzwerks FACEBOOK<sup>7</sup>. Als *Korpus* wird hier eine Sammlung maschinenlesbarer Daten bezeichnet, die unter natürlichen Kommunikationsbedingungen entstanden sind. Die Auswahl der Daten soll repräsentativ und ausgewogen sein, das heißt alle relevanten Aspekte entsprechend ihrer natürlichen Verteilung und Bandbreite berücksichtigen. [PS13, S. 7]

#### 4.1.1 Relevante Profile

Das Zusammenstellen und Annotieren des Datensatzes für einen Goldstandard ist typischerweise ein manueller Prozess, der mit großem zeitlichen und personellen Aufwand verbunden ist. Grundsätzlich kann man davon ausgehen, dass eine größere Menge von Daten bessere Ergebnisse beim Lernen von Klassifizierungsalgorithmen erzielt. Eine allgemeine Empfehlung für den benötigten Umfang eines Korpus kann jedoch nicht gegeben werden, da dies unmittelbar von der konkreten Aufgabenstellung abhängt. Für Anwendungen im Umfeld des *Natural Language Processing*, bei denen rein textuelle Daten verarbeitet werden, wurden bereits Ansätze für die systematische Bestimmung geeigneter Korpusgrößen vorgeschlagen. Häufig jedoch werden Entscheidungen über den Umfang eines Datensatzes durch praktische Erwägungen sowie zeitliche, personelle und finanzielle Ressourcen bestimmt. [Juc12] [PS13, S. 44]

Wichtiger als der Umfang des Datensatzes sind jedoch die *Repräsentativität* und *Vollständigkeit* der Daten. Das heißt die Daten des Goldstandards dürfen nicht willkürlich gewählt sein und sollen die Verteilung der Merkmale der Grundgesamtheit des spezifischen Untersuchungsgegenstandes möglichst genau widerspiegeln sowie die gesamte Bandbreite des Klassifizierungsproblems berücksichtigen. [PS13, S. 7] Für die vorliegende Aufgabe bedeutet das, einen thematisch passenden und repräsentativen Referenzdatensatz für die Detektion von Gefährdern in sozialen Online-Netzwerken zu erstellen. Darüber hinaus müssen die Daten entsprechend gelabelt, das heißt klassifiziert sein, damit diese als Trainings- und Testdaten geeignet sind (siehe Abschnitt 4.2).

#### 4.1.2 Relevante Attribute

Untersuchungen haben gezeigt, dass die Anzahl von Trainingsbeispielen, die notwendig sind, um eine gegebene *Accuracy* zu erreichen, mit zunehmender Anzahl irrelevanter *Features* exponentiell ansteigen kann. Darüber hinaus neigen Lernmodelle bei einer großen Anzahl von *Features* zum *Overfitting* (siehe Abschnitt 3.2.1), was deren Performance tendenziell verschlechtert. Auf Basis weniger und relevanter *Features* können Modelle besser und schneller gelernt werden, bei insgesamt geringeren Anforde-

<sup>7</sup> Zu den Vorteilen von FACEBOOK für die Forschungsarbeit und die besondere Eignung von FACEBOOK Pages siehe Abschnitt 2.4.



rungen an die technischen Ressourcen und besserer Interpretierbarkeit des Lernmodells. [BL97] [Agg14, S. 38ff.]

Schon beim Zusammenstellen der Daten für den Goldstandard ist daher wichtig, möglichst nur Attribute aufzunehmen, bei denen eingeschätzt werden kann, dass diese für die spätere Klassifizierung relevant sind. Die im Rahmen dieser Arbeit betrachteten FACEBOOK *Pages* besitzen eine Vielzahl von Attributen (siehe Abschnitt 2.4.2). Aus allen verfügbaren Attributen müssen für den Goldstandard diejenigen ausgewählt werden, die für eine Differenzierung und Klassifizierung der Profile geeignet erscheinen, und hier als *Feature-Kandidaten* bezeichnet werden.

Ausgehend von diesen Vorüberlegungen wurden folgende Kriterien und Anforderungen für die Auswahl eines Attributs als Feature-Kandidat definiert:

- **Öffentlich**  
Das Attribut muss ohne *Access Token* und ohne explizite Freigabe durch einen Administrator einer *Page* verfügbar sein.
- **Universell**  
Das Attribut soll bei allen Typen von *Pages* und unabhängig von besonderen Anwendungsfällen vorhanden sein (beispielsweise ist das Attribut *food styles* ungeeignet, da es nur bei *Pages* vom Typ *Restaurant* existiert).
- **Messbar**  
Das Attribut soll maschinell gut auswertbar und durch einen Klassifizierungsprozess zu verarbeiten sein (beispielsweise sind die Anforderungen bei Video- und Audiodaten höher als bei rein textuellen Informationen).
- **Diskriminierend**  
Das Attribut soll eine spezifische Eigenschaft einer *Page* beschreiben, die grundsätzlich geeignet scheint, Profile hinsichtlich des Zielkonzeptes (Zielklassen) voneinander abzugrenzen (beispielsweise ist das Attribut *id* ungeeignet, da es nur eine technische Funktion hat).

Aus den gewählten Feature-Kandidaten (siehe Abschnitt 4.3) können in einem späteren Verarbeitungsschritt relevante *Features* selektiert, gegebenenfalls neue *Features* extrahiert und für das Training eines konkreten Klassifikator genutzt werden (siehe Abschnitt 3.2.1).

### 4.1.3 Abgrenzung der Daten

Eine wichtige Anforderung für den Goldstandard ist eine möglichst *balancierte Verteilung* der Daten sowohl in Bezug auf die Zielklassen als auch hinsichtlich des Umfangs der je Trainingsbeispiel berücksichtigten Daten. [BL97] [PS13, S. 7f.]

**Herkunft** Typischerweise sind mit den *Edges* einer FACEBOOK *Page* viele *Nodes* unterschiedlichen Typs verbunden, deren Inhalte unterschiedlicher Herkunft sein können (siehe Abschnitt 2.4.2). Neben eigenen *Posts*, *Photos* und *Comments*, die durch den oder die Betreiber einer *Page* veröffentlicht wurden, gibt es auch Inhalte, die von Besuchern der *Page* publiziert wurden. Für die Klassifizierung eines Profils und damit auch für das Training eines Klassifikators sind in Bezug auf den Goldstandard jedoch nur die Inhalte relevant, die durch eine *Page* originär veröffentlicht wurden.

**Reichweite** Die starke Vernetzung von Akteuren und publizierten Inhalten ist ein exponiertes Merkmal sozialer Online-Netzwerke. Nahezu alle (für einen bestimmten Nutzer sichtbaren) Inhalte anderer Profile können geteilt, bewertet und kommentiert werden (siehe Abschnitt 2.3). Werden nun beispielsweise beliebige *Posts* oder *Comments*, welche durch eine betrachtete FACEBOOK *Page* auch auf anderen Profilen erstellt wurden, aus dem Netzwerk gefiltert, so geht deren Kontext in denjenigen Fällen verloren, in denen der ursprüngliche Beitrag von einem anderen Profil stammte. Für den Aufbau des Goldstandards sind nach Einschätzung des Autors jedoch nur diejenigen Inhalte und Aktivitäten einer FACEBOOK *Page* relevant, deren Kontext unter Berücksichtigung der oben beschriebenen Herkunft erhalten bleibt. Das schließt unter anderem *Comments* aus, die von einer betrachteten FACEBOOK *Page* in Reaktion auf *Comments* anderer Profile zu ihren eigenen Inhalten publiziert wurden.

**Volumen** Je nach Aktivität der Betreiber eines FACEBOOK-Profiles kann beispielsweise die Anzahl von *Posts* und *Comments* in der Spanne von einigen wenigen im Jahr bis hin zu mehreren pro Tag liegen. Um das Training des Lernmodells durch den Einfluss einzelner Profile mit großer Aktivität nicht zu verzerren, sollte der Umfang der je Profil berücksichtigten Daten über alle Trainingsbeispiele hinweg weitgehend gleich verteilt sein. Theoretisch denkbar wären eine Begrenzung des betrachteten Datenvolumens je Attributtyp (beispielsweise eine Anzahl Kilobyte oder Wörter an Text) oder eine definierte Anzahl von Einzelobjekten je Attributtyp (beispielsweise 10 *Posts*, 10 *Comments* und 10 *Photos*).

Zusammenfassend schlägt der Autor für den im Rahmen dieser Arbeit entwickelten Goldstandard vor, diejenigen Inhalte einer FACEBOOK *Page* zu berücksichtigen, die durch einen Besucher auf FACEBOOK visuell als Bestandteil dieser *Page* wahrgenommen werden (siehe Abbildung 4.1) und durch den Betreiber dieser *Page* selbst erstellt wurden. Bezogen auf den FACEBOOK *Social Graph* umfasst dies die **Inhalte aller relevanten Fields derjenigen Nodes**, die **innerhalb der 2-Schritt-Umgebung** einer *Page* liegen und deren **Urheber die Page** selbst ist (siehe Abschnitt 2.4.2). Um Profile mit unterschiedlicher Aktivität auszuwogen einzubeziehen, sollten jeweils die **20 jüngsten** (und damit aktuell repräsentativsten) **Objekte** jeden *Node*-Typs (*Posts*, *Comments*, *Photos* usw.) aufgenommen werden. Dieser Vorschlag ist als initiale Konfiguration zu verstehen, deren Angemessenheit und Eignung als Goldstandard in späteren Arbeiten evaluiert und welche daraufhin gegebenenfalls überarbeitet werden sollte.

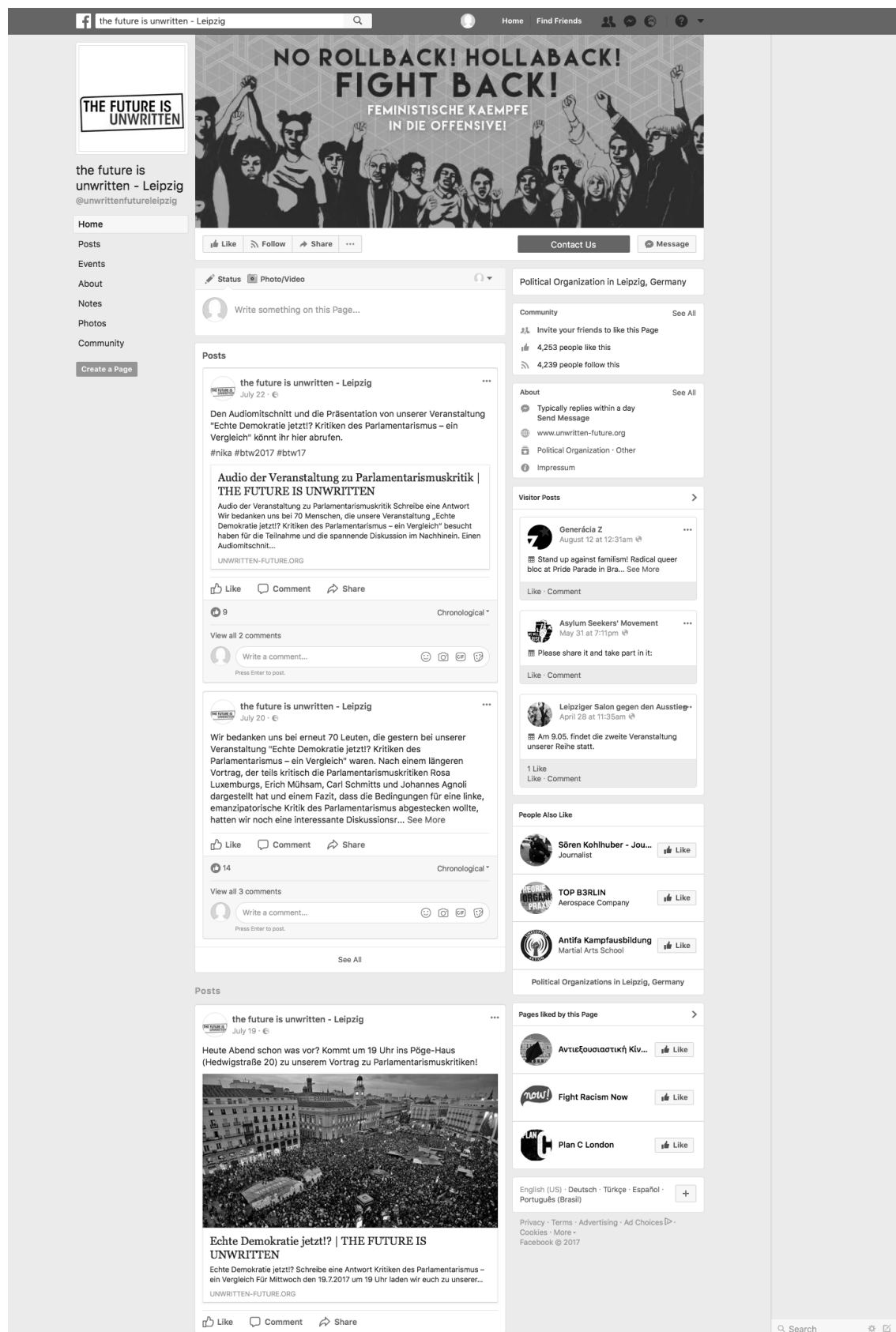


Abbildung 4.1: Exemplarischer Screenshot mit den Elementen einer FACEBOOK Page

#### 4.1.4 Bereitstellung der Daten

Die Daten für den Goldstandard müssen so bereitgestellt werden, dass diese möglichst einfach und universell für das Training und die Evaluation von Klassifikatoren in unterschiedlichen Umgebungen genutzt werden können. Daraus ergeben sich einige spezielle Anforderungen an das Format und die technische Ablage der Daten (vgl. [CDRT07]):

- **Unabhängigkeit** von konkreten Herstellern oder Applikationen,
- **Interoperabilität** zwischen Applikationen, aber auch Organisationen,
- **Erweiterbarkeit** für wechselnde Anforderungen und neue Aufgabenstellungen,
- **Einfacher** maschineller **Zugriff** und **einfache** maschinelle **Verarbeitung**,
- **Skalierbarkeit** für große Mengen von Profilen oder Features,
- **Robustheit** gegenüber fehlenden oder fehlerhaften Werten.

Bei vielen existierenden Spezifikationen für die Beschreibung von Daten aus sozialen (Online-)Netzwerken liegt der Fokus auf der statischen Abbildung von (menschlichen) Akteuren und den (expliziten) Strukturen eines Netzwerkes. Interaktionen und publizierte Inhalte als herausragende Merkmale der Dynamik sozialer Online-Netzwerke werden kaum berücksichtigt. Häufig werden die Daten mit einer flachen Struktur und mit kumulierten Häufigkeitswerten für das Auftreten bestimmter Typen von Beziehungen oder Eigenschaften in einfach strukturierten Textdateien (beispielsweise *CSV – Comma Separated Values*) abgelegt (vgl. [MyP15]).

Es gibt einige Spezifikationen, die auf **RDF (Resource Description Framework)** basieren, einem Standard zum Beschreiben von Online-Ressourcen beliebiger Art im *Semantic Web* [TAH<sup>+</sup>16, S. 62]. Die RDF-Schemata *FOAF (Friend Of A Friend)* und *SIOC (Semantically-Interlinked Online Communities)* beispielsweise definieren *Ontologien* für die Beschreibung persönlicher Daten und Beziehungen zwischen Menschen und deren Rollen in sozialen Netzwerken respektive für die Abbildung von *Communities* im *Social Web* [Ro09]. *SemSNA (Semantic Social Network Analysis)* – eine ebenfalls RDF-basierte Ontologie zur Abbildung der spezifischen Metriken und zur Analyse sozialer Netzwerke – wiederum hat das Ziel, die Strukturen sozialer Netzwerke von deren konkreter Domäne zu abstrahieren und mit Annotationen anzureichern [EBGC09]. Die genannten Spezifikationen sind durch ihren generischen Ansatz der Beschreibung von Objekten bezogen auf ihren konzeptuellen Fokus recht flexibel, dadurch aber schlechter lesbar für menschliche Nutzer und erfordern bei der maschinellen Verarbeitung ein aufwändiges Mapping der generischen Datenstrukturen auf die spezifische Struktur einer Anwendung. Darüber hinaus können besondere Merkmale sozialer Online-Netzwerke, beispielsweise Interaktionen wie Teilen, Bewerten und Kommentieren von Inhalten anderer Nutzer, nicht explizit abgebildet werden. [BR14]

Ein ebenfalls häufig genutzter Ansatz zur Beschreibung von Daten im Umfeld der Analyse sozialer (Online-)Netzwerke ist **XML (*Extensible Markup Language*)**. XML ist eine vom *W3C (World Wide Web Consortium)* standardisierte Auszeichnungssprache zur strukturierten Beschreibung von Daten und Dokumenten in Form von Textdateien, die sich besonders für den Austausch von Daten und Dokumenten im Internet eignet. Mittels einer *DTD (Document Type Definition)* oder mit Schema-Sprachen wie *XML Schema* können die Struktur und Grammatik anwendungsspezifischer Sprachen auf Basis von XML definiert werden [TAH<sup>+</sup>16, S. 62] (vgl. [Wor16a]). *DyNetML* ist beispielsweise ein XML-Schema, dessen Fokus auf der sogenannten dynamischen Netzwerkanalyse liegt und mit dem der Graph eines sozialen Netzwerks abstrakt modelliert werden kann, um die Veränderungen von Beziehungen im Zeitverlauf zu beschreiben und zu analysieren. Auch hier werden ausschließlich die Beziehungen zwischen den Knoten betrachtet und spezifische Eigenheiten der Interaktionen sozialer Online-Netzwerke lassen sich nicht abbilden (vgl. [CDRT07]).

XML als Beschreibungssprache bietet jedoch einige Vorteile und eignet sich grundsätzlich als universelles Ablage- und Austauschformat von Daten [CDRT07] [PS13, S. 82f.]:

- **ISO- und Industriestandard** für die strukturierte Beschreibung und den Austausch von Daten,
- **Plattformunabhängig** hinsichtlich Rechnerarchitektur, Betriebssystem und Programmiersprache,
- **Gut verfügbare Infrastruktur** für Spezifikation, Zugriff und Verarbeitung (Editoren, Parser, Programmierschnittstellen, Datenbanken),
- Hierarchische Struktur von XML-Dokumenten **gut geeignet für die Abbildung von Teilgraphen** eines Netzwerks,
- **Einfaches Mapping** auf anwendungsspezifische Strukturen,
- **Gut transformierbar** in andere XML-Sprachen, -Strukturen und Datenformate mittels *XSLT (Extensible Stylesheet Language Transformation)*,
- **Für Menschen gut lesbare** Textdateien,
- Wegen der expliziten und eindeutigen Struktur **gut maschinenlesbar**,
- **Flexibel und gut erweiterbar**, ohne Abwärtskompatibilität zu existierenden Anwendungen zu verlieren (zusätzliche Annotationen, neue Akteure, neue Features, neue soziale Netzwerke).

Ausgehend von diesen Vorüberlegungen und den genannten Vorteilen von XML für die Beschreibung strukturierter Daten schlägt der Autor im Rahmen dieser Arbeit die Entwicklung eines eigenen XML-Schemas für die Abbildung der Daten des Goldstandards vor (siehe Abschnitt 4.4).

## 4.2 Auswahl geeigneter Profile

Das SÄCHSISCHE LANDESAMT FÜR VERFASSUNGSSCHUTZ (LFV) hat 2016 einen Bericht über die Aktivitäten sächsischer extremistischer Gruppierungen im Internet veröffentlicht. In diesem Bericht sind unter anderem die Profile der beobachteten links- und rechtsextremistischen Gruppierungen in sozialen Online-Netzwerken aufgeführt, sofern vorhanden und aktiv genutzt (Stand Juni 2016, vgl. [Lan16]).

Zentrale Aufgabe der LANDESÄMTER FÜR VERFASSUNGSSCHUTZ in Deutschland ist der Schutz der „freiheitlich-demokratischen Grundordnung“, unter anderem durch die Beobachtung extremistischer Personen und Gruppierungen, um sich über deren Vorhaben und Ziele bereits im Vorfeld von Straftaten ein Bild zu machen. [Lan17a, S. 289ff.]

Auch wenn die Bundesregierung ausdrücklich betont, dass der Begriff des *Gefährders* der polizeilichen Terminologie entstammt und im Bereich des Verfassungsschutzes keine Anwendung findet, weisen die Aspekte „extremistische Personen und Gruppierungen“ sowie „im Vorfeld von Straftaten“ darauf hin, dass sich sowohl für die Einstufung von Gefährdern durch die Polizeibehörden als auch für die Beobachtung durch den Verfassungsschutz vergleichbare Ziele, Voraussetzungen und Kriterien für eine entsprechende Beurteilung relevanter Personen und Gruppierungen ableiten lassen (vgl. auch Abschnitt 3.1). [Deu17a, Deu17b]

Obwohl das Vorgehen und die konkreten Kriterien für die Einstufung und Beobachtung einer extremistischen Gruppierung nicht bekannt sind, kann auf Basis der vorliegenden Informationen davon ausgegangen werden, dass der Datensatz des SÄCHSISCHEN LANDESAMTES FÜR VERFASSUNGSSCHUTZ über extremistische Gruppierungen in Sachsen die in Abschnitt 4.1.1 beschriebenen Anforderungen an die Repräsentativität für die Auswahl relevanter Profile überwiegend erfüllt. Weiterhin ist dieser Datensatz besonders qualifiziert, da er in Bezug auf den Untersuchungsgegenstand nicht nur repräsentativ, sondern darüber hinaus mit den notwendigen Labels für die Zielklassen versehen ist. Damit entfällt die sonst notwendige, aufwändige manuelle Annotation der Referenzdaten, da diese bereits im Vorfeld durch die Experten des SÄCHSISCHEN LANDESAMTES FÜR VERFASSUNGSSCHUTZ vorgenommen wurde (vgl. [Lan16]).

Profile extremistischer Gruppierungen in sozialen Online-Netzwerken werden aus verschiedenen Gründen gelegentlich sowohl seitens der Strafverfolgungsbehörden, aber auch durch die Betreiber selbst, geschlossen und teilweise durch ein neues Profil beim gleichen oder einem anderen sozialen Online-Netzwerk ersetzt. Für die Auswahl relevanter FACEBOOK-Profilen wurde deshalb zunächst geprüft, ob diese zum Untersuchungszeitpunkt noch online verfügbar sind (Stand September 2017). Falls das im Bericht genannte Profil nicht mehr erreichbar war, wurde weiter untersucht, ob die betreffende Gruppierung zwischenzeitlich ein neues Profil bei FACEBOOK eröffnet hat. Dieses wurde ersatzweise jedoch nur dann herangezogen, wenn es eindeutig der betreffenden

Gruppierung zugeordnet werden konnte (Prüfung der *Fields Name*, *Description* und *About*).

Die Auswertung der extremistischen Gruppierungen aus dem Bericht des LFV ergab nach Bereinigung der nicht mehr existierenden FACEBOOK-Profile folgende Verteilung. Eine detaillierte Aufstellung findet sich im Anhang C:

- **16 linksextremistische Profile**, die durch das LFV als *autonome Gruppierung*, *anarchistische Gruppierung*, *Partei*, *Musik* oder *Publikation* typisiert wurden.
- **57 rechtsextremistische Profile**, die durch das LFV als *Partei*, *Subkultur*, *Musik*, *Geschäft*, *Event* oder *Sonstige* typisiert wurden.

Sowohl die links- als auch die rechtsextremistischen Profile können als *Gefährder* im Sinne dieses Goldstandards interpretiert werden. Im Rahmen der späteren Entwicklung und des Trainings von Klassifikatoren müsste geprüft werden, ob sich gemeinsame Indikatoren für eine Klassifizierung der links- und rechtsextremistischen Profile als *Gefährder* finden oder ob die beiden Klassen bei der zukünftigen Arbeit an Algorithmen für die Detektion von Gefährdern weiterhin getrennt betrachtet werden.

Das Training von Klassifikatoren für die Detektion von Gefährdern in sozialen Online-Netzwerken benötigt jedoch nicht nur Trainingsbeispiele für die Klasse der *Gefährder*, sondern auch Beispiele für die Klasse der *Nicht-Gefährder*. Im Umkehrschluss zur Einstufung von Profilen als *extremistisch* müssten theoretisch alle übrigen Profile als *nicht-extremistisch* bewertet werden und damit als neutrale Trainingsbeispiele geeignet sein. Da es im Gegensatz zur Klasse der *Gefährder* keine zentrale Instanz für die Bewertung von Gruppen oder Personen als *Nicht-Gefährder* gibt, muss die Auswahl und Annotation entsprechender Profile aus dieser Grundgesamtheit der nicht als *Gefährder* eingestuft Profile jedoch manuell durch Experten erfolgen.

Als Ausgangspunkt für die Selektion geeigneter Profile könnten staatliche Institutionen dienen, die gesetzlich zur Neutralität verpflichtet sind, beispielsweise Einrichtungen der staatlichen Verwaltung, des Justizwesens, der öffentlichen Bildung, der gesetzlichen Krankenkassen, des öffentlichen Personenverkehrs und der öffentlich-rechtlichen Medien. Darüber hinaus gibt es auch nicht-staatliche Bereiche, bei denen von einer faktischen Neutralität ausgegangen werden kann, weil diese als Zielgruppe für sich die Gesamtheit oder überwiegende Mehrheit aller Menschen beanspruchen. Dazu gehören beispielsweise Marken und Produkte großer Hersteller, aber auch Akteure aus den Bereichen Tourismus und Sport. Im zeitlichen Rahmen der vorliegenden Arbeit war es dem Autor selbst leider nicht möglich, eine solche Auswahl neutraler Profile zu erstellen, so dass der Goldstandard in einem zukünftigen Arbeitsschritt um diese Trainingsbeispiele ergänzt werden muss.

## 4.3 Auswahl von Feature-Kandidaten

In Abschnitt 4.1.2 wurden Anforderungen für die Auswahl der Feature-Kandidaten für den Goldstandard definiert. Auf Basis dieser Kriterien und anhand der Beschreibung der Attribute in der *FACEBOOK Graph API Documentation* wurden die einzelnen *Fields* der relevanten *Nodes* bewertet und ein Katalog von Feature-Kandidaten erstellt. Dabei wurden auch – wie in Abschnitt 4.1.3 beschrieben – die Herkunft und Reichweite der Daten innerhalb der 2-Schritt-Umgebung einer *FACEBOOK Page* berücksichtigt (vgl. auch Abschnitt 2.4.2 und Abbildung 2.2). Auf Basis des Feature-Katalogs müssen dann im weiteren Prozess durch *Feature Selection* und *Extraction* geeignete *Features* für das konkrete Verfahren und das konkrete Lernmodell gefunden werden (siehe Abschnitt 3.2.1).

Die Aufstellung des Feature-Katalogs erfolgte mehrstufig in folgenden Schritten:

1. **Auswahl der allgemein verfügbaren Attribute der relevanten *Nodes***

Nur diejenigen *Fields* und *Edges*, die nicht als *special*, *restricted* oder *deprecated* markiert sind (vgl. auch Abschnitt 2.4.2).

2. **Auswahl der relevanten Feature-Kandidaten**

Gemäß dem Anforderungskatalog aus Abschnitt 4.1.2 relevante *Fields* derjenigen *Nodes*, die innerhalb der *2-Schritt-Umgebung* einer *Page* liegen und deren Urheber die *Page* selbst ist (vgl. Abschnitt 4.1.3).

3. **Ergänzen sinnvoller Meta-Informationen**

*Fields*, deren Inhalte für die Identifizierung, Ermittlung der Herkunft oder weitere Verarbeitung der Daten sinnvoll sein könnten, beispielsweise Objekt-IDs und kumulierte Anzahlen von *Likes* und *Comments* als Maß für die Popularität eines Objektes.

Der komplette Katalog von Feature-Kandidaten für die Entitäten *Album Node*, *Comment Node*, *Event Node*, *Page Node*, *Photo Node*, *Picture Node*, *Post Node*, *User Node*, *Video Node* und *VideoList Node* ist in Anhang B zusammengestellt. In den Tabellen im Anhang sind jeweils alle allgemein verfügbaren Attribute dargestellt (siehe Schritt 1). Die für den Goldstandard ausgewählten Attribute sind mit *considered* bzw. *partially considered* markiert. *Partially considered* bedeutet, dass nur derjenige Teil der verbundenen *Nodes* relevant ist, bei dem die Inhalte durch die *Page* selbst publiziert wurden (siehe Schritt 2). Die im Schritt 3 zusätzlich gewählten Attribute mit Meta-Informationen sind mit *meta* markiert.

Beispielhaft sind in den Tabellen 4.1 und 4.2 die Feature-Kandidaten eines *Page Node* aufgeführt<sup>8</sup>.

<sup>8</sup> Aus Platzgründen sind hier nur die tatsächlich berücksichtigten Attribute dargestellt.



Tabelle 4.1: Feature-Kandidaten der *Fields* eines *Page Node* (nach [Fac17e])

Field	Type	Description	Note
about	string	Information about the Page	considered
best_page	Page	The best available Page on Facebook for the concept represented by this Page.	meta
category	string	The Page's category. e.g. Product/Service, Computers/Technology	considered
category_list	list<PageCategory>	The Page's sub-categories	considered
description	string	The description of the Page	considered
fan_count	unsigned int32	The number of users who like the Page. For Global Pages this is the count for all Pages across the brand.	meta
featured_video	Video	Video featured by the Page	considered
general_info	string	General information provided by the Page	considered
id	numeric string	Page ID. No access token is required to access this field	meta
impressum	string	Legal information about the Page publishers	considered
is_community_page	bool	Indicates whether the Page is a community Page	meta
link	string	The Page's Facebook URL	meta
name	string	The name of the Page	considered
overall_star_rating	float	Overall page rating based on rating survey from users on a scale of 1-5. This value is normalized and is not guaranteed to be a strict average of user ratings.	meta
rating_count	unsigned int32	Number of ratings for the page.	meta
talking_about_count	unsigned int32	The number of people talking about this Page	meta
username	string	The alias of the Page. For example, for <a href="https://www.facebook.com/platform">www.facebook.com/platform</a> the username is 'platform'	considered
verification_status	string	Showing whether this Page is verified and in what color e.g. blue verified, gray verified or not verified	meta
website	string	The URL of the Page's website	meta

Tabelle 4.2: Feature-Kandidaten der *Edges* eines *Page Node* (nach [Fac17e])

Edge	Type	Description	Note
albums	list<Album>	Photo albums for this Page	considered
events	list<Event>	Events for this Page	considered
photos	list<Photo>	This Page's photos	considered
picture	Picture	This Page's profile picture	considered
posts	Post[]	The feed of posts (including status updates) and links published by this page, or by others on this page.	considered
video_lists	list<VideoList>	Video Playlists for this Page	considered
videos	list<Video>	Videos for this Page	considered

## 4.4 Struktur und Format der Daten

Die Daten des Goldstandards werden dem entwickelten Vorschlag folgend (siehe Abschnitt 4.1.4) in einer XML-Struktur abgelegt. Um einen einfachen Austausch zwischen verschiedenen Werkzeugen oder Partnern zu ermöglichen, sollen die Daten jedes Profils als separate XML-Textdatei im Dateisystem gespeichert werden. Der Dateiname der XML-Datei entspricht der FACEBOOK *ID* des Profils. Die in einem Profil referenzierten Medienobjekte (Fotos, Videos, Audio-Dateien) werden ebenfalls im Dateisystem in einem separaten Ordner je Profil mit ihrer jeweiligen FACEBOOK *ID* als Dateiname abgelegt. Das Layout der Dateien und Ordner im Dateisystem stellt sich exemplarisch wie folgt dar:

```

goldstandard
├── xml
│   └── <facebookPageId>.xml
├── media
│   ├── <facebookPageId>
│   │   ├── audio
│   │   │   └── <facebookAudioId>.<ext>
│   │   ├── photo
│   │   │   └── <facebookPhotoId>.<ext>
│   │   └── video
│   │       └── <facebookVideoId>.<ext>
└── xsd
    └── facebookgoldstandard.xsd

```

Die im obigen Dateilayout aufgeführte Datei *facebookgoldstandard.xsd* enthält die *XML Schema Definition (XSD)*, die Spezifikation des sogenannten XML-Schemas für die XML-Dateien des Goldstandards. Die *XSD* definiert die Strukturen und Datentypen einer im Sinne dieses Schemas gültigen XML-Instanz. Durch die gemeinsame Nutzung

eines XML-Schemas für den Austausch und die Verarbeitung von XML-Dateien können die korrekte *Struktur* und *Semantik* über alle beteiligten Anwendungen und Partner hinweg sichergestellt werden. Die Modellierung des XML-Schemas erfolgte mit den XML-TOOLS der ECLIPSE FOUNDATION [Ecl16, Ecl17]; für die grafische Visualisierung wurde der ONLINE XML SCHEMA VIEWER von RAFFELBERGER genutzt [Raf17].

Mit *XML Schema* können auch komplexe Datentypen benannt und global deklariert werden, was generell die Lesbarkeit und Wartbarkeit und vor allem die Wiederverwendbarkeit und Erweiterbarkeit des Schemas erleichtert (zu den Anforderungen der Datenablage siehe Abschnitt 4.1.4). Die globalen Datentypen lassen sich für die weitere Verarbeitung auch gut auf Klassen in objektorientierten Programmierumgebungen oder Datenbanken abbilden. Eine *XSD* bietet darüber hinaus die Möglichkeit, „fremde“ Schema-Spezifikationen zu importieren, was beispielsweise die spätere Erweiterung des Goldstandards um Strukturen für andere soziale Online-Netzwerke unter Verwendung der bereits deklarierten Datentypen ermöglicht. [Wor12a, Wor12b, Wor16b]

Das XML-Schema definiert das Element *page* als Wurzelement einer gültigen XML-Instanz. Dieses Element besitzt XML-Attribute, welche die als relevant definierten *Fields* einer FACEBOOK *Page* abbilden. Weiterhin enthält das *page*-Element in einer hierarchischen Struktur alle weiteren Elemente, welche die relevanten Eigenschaften derjenigen *Nodes* repräsentieren, die innerhalb der 2-Schritt-Umgebung als Teilgraph mit der *Page* verbunden sind (vgl. Abschnitt 4.3).

Für jeden *Node*-Typ wurde ein separater globaler Datentyp deklariert, welcher die relevanten *Fields* und *Edges* dieses Typs modelliert (*albumType*, *commentType*, *eventType*, *pageType*, *photoType*, *pictureType*, *postType*, *userType*, *videoType*, *videoListType*). Weiterhin wurde für die Abbildung der *Edges* jeweils ein eigener globaler Datentyp deklariert, welcher als Containerelement die über eine *Edge* verbundenen *Nodes* enthält (*albumsType*, *commentsType*, *eventsType*, *photosType*, *postsType*, *videoListsType*, *videosType*).

Ergänzend zu den Elementen und Attributen, welche die relevanten *Fields* und *Edges* der Objekte aus FACEBOOK abbilden, wurden im XML-Schema für den Goldstandard für jeden *Node*-Typ zusätzliche Attribute vorgeschlagen, welche vorgegebene (*a priori*) und berechnete (*calculated*) *Label*- und *Score*-Werte speichern können. Diese zusätzlichen Attribute wurden sowohl dem Element *page* als auch allen anderen Elementen – welche einen *Node*-Typ repräsentieren – hinzugefügt, um auch Zwischenschritte und -ergebnisse der in Abschnitt 3.2.3 beschriebenen Klassifikatoren-Ensembles abbilden zu können.

Der vollständige Quelltext der XML-Schema-Datei ist in Anhang D abgebildet. Aufbau und Struktur des XML-Schemas sind in den Abbildungen 4.2 und 4.3 zu sehen.

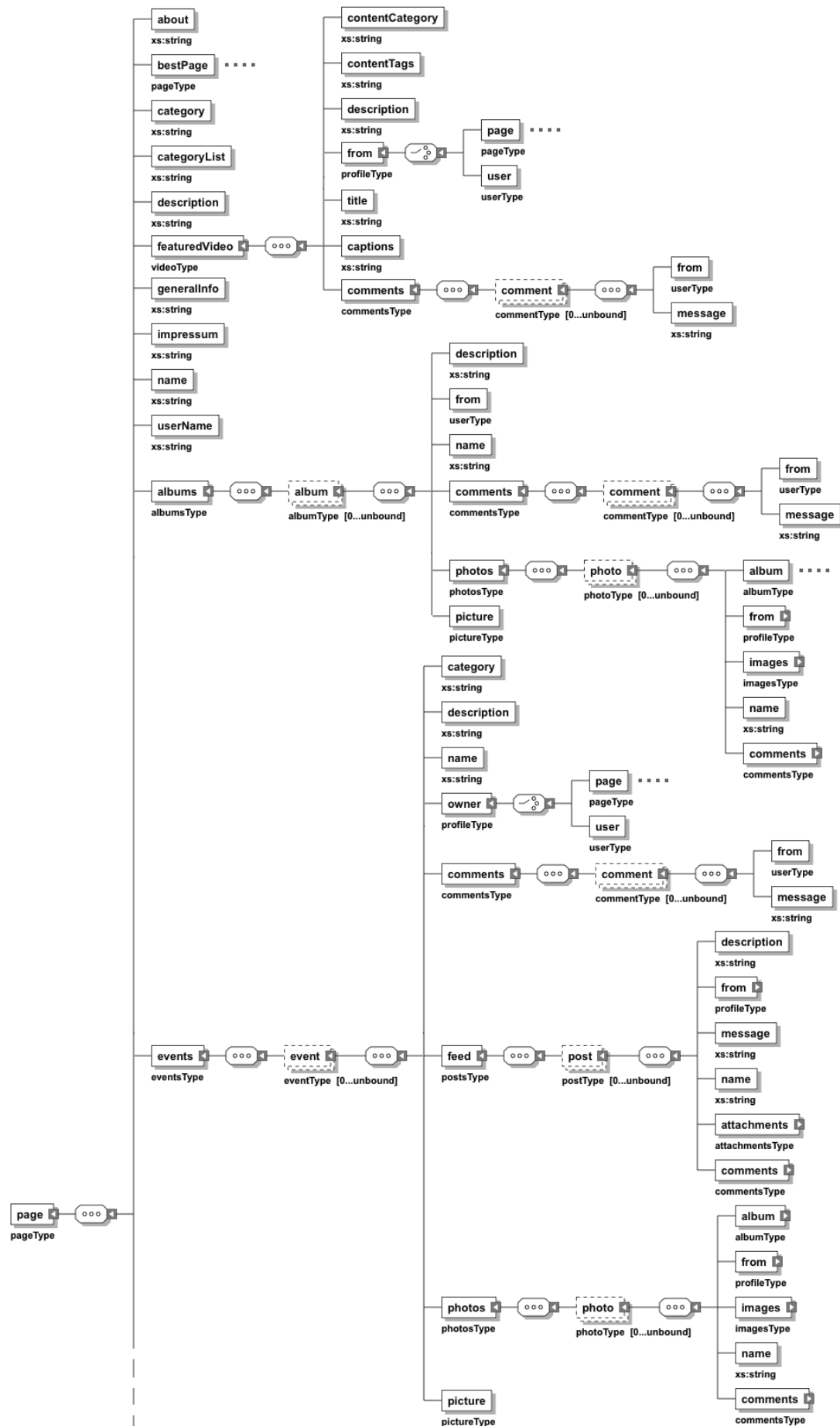


Abbildung 4.2: Struktur des XML-Schemas für die Abbildung der Profildaten des Goldstandards (Fortsetzung auf folgender Seite)

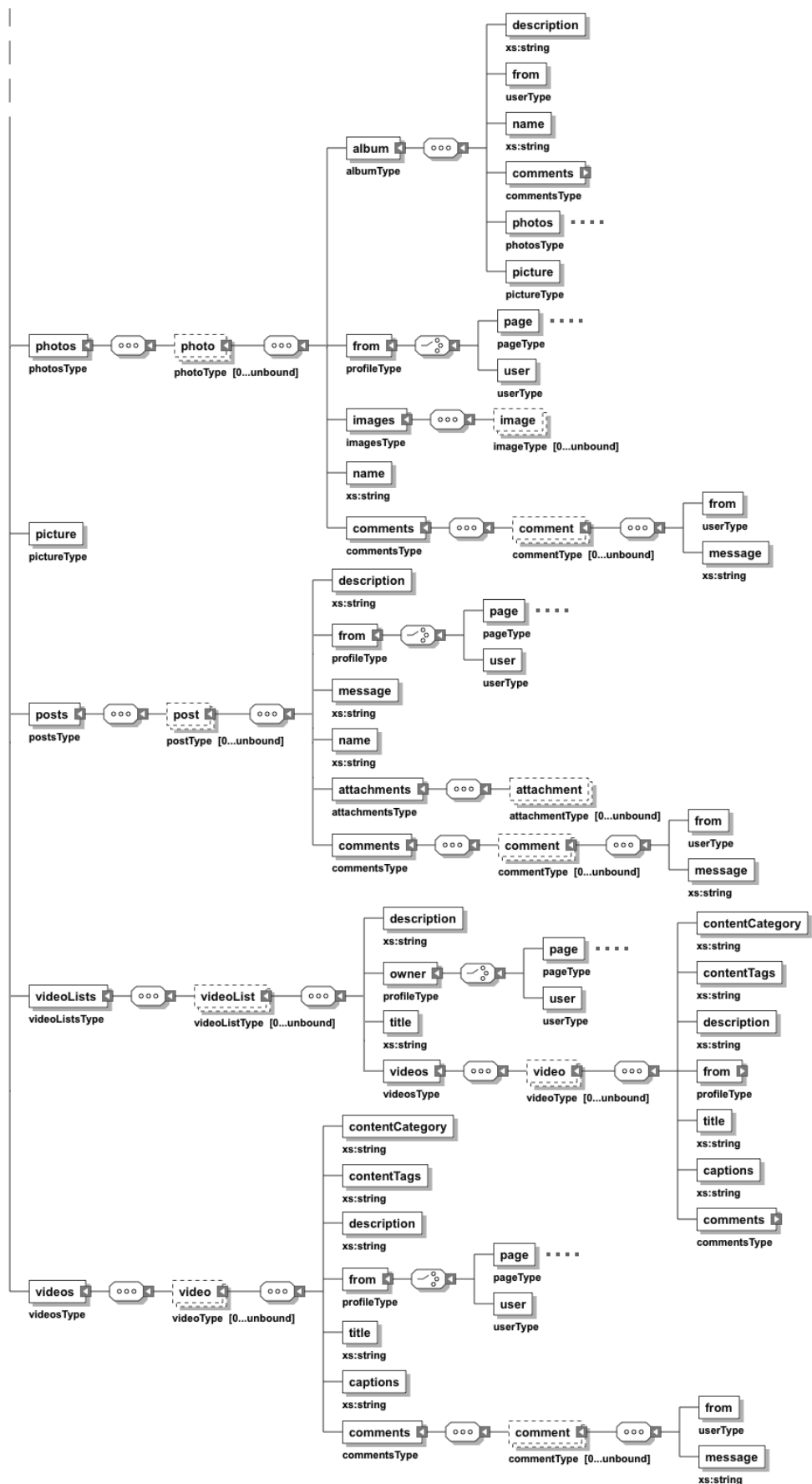


Abbildung 4.3: Struktur des XML-Schemas für die Abbildung der Profildaten des Goldstandards (Fortsetzung von vorheriger Seite)



## 5 Fazit und Ausblick

Im Ergebnis der vorliegenden Arbeit wurde die **Basis für einen Goldstandard** geschaffen, auf dessen Grundlage Algorithmen zur Detektion von Gefährdern in sozialen Online-Netzwerken entwickelt, trainiert und evaluiert werden können. Zunächst wurden die besonderen Eigenschaften sowie die **Modellierung und Analyse sozialer Online-Netzwerke** als spezielle Ausprägung sozialer Netzwerke betrachtet und die Eignung des sozialen Online-Netzwerks Facebook als Forschungsgegenstand untersucht (vgl. Kapitel 2). Anschließend wurden **spezifische Merkmale von Gefährdern** identifiziert und die besonderen Herausforderungen bei der Anwendung von **Klassifizierungsverfahren auf Daten aus Social-Media-Quellen** herausgearbeitet (vgl. Kapitel 3). Ausgehend von diesen Vorüberlegungen wurden **Anforderungen für die Realisierung des Goldstandards** definiert (vgl. Kapitel 4). Diese Anforderungen umfassen neben Kriterien für die Auswahl relevanter Profile und Attribute auch Vorschläge für die Abgrenzung von Herkunft, Reichweite und Volumen der zu berücksichtigenden Daten eines FACEBOOK-Profiles. Weiterhin wurden detaillierte **Vorgaben für die Bereitstellung der Daten** des Goldstandards formuliert. Ausgehend von diesen Anforderungen und Rahmenbedingungen wurden **konkrete Vorschläge für einen Datensatz** sowohl hinsichtlich **geeigneter FACEBOOK-Profile** als auch **entsprechender Feature-Kandidaten** erarbeitet. Für die Strukturierung und Bereitstellung der Daten wurden ein **anwendungsbereites XML-Schema** und ein **Layout zur Ablage der Daten** des Goldstandards im Dateisystem entwickelt. Die Sammlung relevanter Profile für den Goldstandard beinhaltet bereits **konkrete Trainingsbeispiele für die Klasse der Gefährder**. Im zeitlichen Rahmen der vorliegenden Arbeit war es leider nicht möglich, auch für die Klasse der *Nicht-Gefährder* konkrete Profile zu selektieren. Es wurde jedoch ein Vorschlag für das **Vorgehen zur Auswahl geeigneter Profile für die Klasse der Nicht-Gefährder** erarbeitet, um den Goldstandard in einem zukünftigen Arbeitsschritt um diese Trainingsbeispiele zu ergänzen.

Ausgehend von den Ergebnissen dieser Arbeit besteht die Aufgabe in einer folgenden Phase darin, ein Werkzeug zu entwickeln, mit welchem der Goldstandard auf Basis der formulierten Anforderungen mit den konkreten Daten der relevanten Profile und verbundenen Entitäten gefüllt werden kann. Bei Versuchen im Rahmen des vorgelagerten Praxismoduls haben sich bei der experimentellen Abfrage der FACEBOOK *Graph API* und beim Speichern der Profildaten verschiedene **Herausforderungen hinsichtlich der Robustheit und Performance** der Kommunikation gezeigt, welche im Zuge der folgenden Entwicklungsarbeiten adressiert werden müssen:

- Unterbrechung von Netzwerkverbindungen auf Grund der langen Sitzungsdauer beim Herunterladen umfangreicher bzw. zahlreicher Profile,
- Abbruch der Verbindung wegen Beschränkungen der FACEBOOK *Graph API* hin-

sichtlich der Anzahl von *API*-Zugriffen innerhalb einer bestimmten Zeitspanne (vgl. [Fac17f]),

- Lange Ladezeiten durch sequenzielles Abarbeiten der Warteschlange herunterzuladender Profile,
- Hohe Zugriffs- und Reaktionszeiten des genutzten *Backends* (ECLIPSE EMFSTORE) bei komplexen und umfangreichen Modellen,
- Speicherüberlauf bei der Darstellung komplexer Modelle mit zahlreichen Entitäten als Graph.

Wenn die Daten des Goldstandards vollständig vorliegen, kann mit **Entwurf, Training und Evaluation von Klassifikatoren** für die verschiedenen Medientypen begonnen werden. Mit Methoden der *Feature Selection* und *Feature Extraction* können aus den vorgeschlagenen *Feature*-Kandidaten für die Klassifizierung passende *Features* gewonnen werden. Anschließend müssen die einzelnen Klassifikatoren mit einer zweckmäßigen Kombination und Wichtung sowie einem geeigneten *Voting*-Verfahren zu einem *Ensemble* zusammengefügt werden. Im Rahmen dieser Arbeiten ist auch zu prüfen, ob die vorgeschlagenen Profile und Attribute sowie die Abgrenzung der Daten hinsichtlich Herkunft, Reichweite und Volumen die formulierten Anforderungen im Hinblick auf die Eignung als Goldstandard erfüllen können oder gegebenenfalls Anpassungen notwendig sind.

Bei der Bearbeitung des vorliegenden Themas haben sich darüber hinaus einige Aspekte ergeben, die **Gegenstand zukünftiger Arbeiten** sein könnten:

- Prüfen links- und rechtsextremistischer Profile auf gemeinsame Indikatoren, die für eine einheitliche Bewertung als *Gefährder* herangezogen werden können,
- Größerer Umfang des Goldstandards (beispielsweise Profile aus ganz Deutschland oder Profile aus weiteren sozialen Online-Netzwerken),
- Detektion von wichtigen Knoten wie Meinungsführer und Multiplikatoren sowie Communities im Netzwerk auf Basis verdächtiger Profile,
- Identifikation von Profilen über die Grenzen sozialer Online-Netzwerke hinweg, um deren Informationen miteinander zu verknüpfen (vgl. [ZL13, SYBT15]),
- Kombination der Klassifizierung mit anderen Methoden des Maschinellen Lernens wie *Named Entity Recognition*, *Topic Mining* und *Sentiment Analysis*,
- Prüfen standardisierter RDF-Spezifikationen wie *SIOC* und *FOAF* hinsichtlich ihrer Eignung als Ontologien für die Instanziierung gefundener Netzwerke.

Abschließend sei noch einmal auf das Eingangszitat von EDWARD SNOWDEN verwiesen: „I do not want to live in a world where everything I do and say is recorded.“ [Gua13] Dieses Zitat vermittelt die dringende Botschaft, bei der intensiven Diskussion konzeptueller und technischer Fragen einer vernetzten Welt den kontinuierlichen Abgleich mit ethisch-moralischen und juristischen Aspekten unserer freiheitlich-demokratischen Gesellschaft nicht aus den Augen zu verlieren.



## Anhang A: Fields und Edges relevanter Facebook Nodes

### A.1 Album Node

Tabelle A.1: *Fields eines Album Node* (nach [Fac17e])

Field	Type	Description	Note
cover_photo	Photo	The ID of the album's cover photo.	
created_time	datetime	The time the album was initially created.	
description	string	The description of the album.	
event	Event	The event associated with this album.	
from	User	The profile that created the album.	
id	string	The album ID.	
link	string	A link to this album on Facebook.	
location	string	The textual location of the album.	
name	string	The title of the album.	
place	Page	The place associated with this album.	
privacy	string	The privacy settings for the album.	
type	enum {app, cover, profile, mobile, wall, normal, album}	The type of the album.	
updated_time	datetime	The last time the album was updated.	

Tabelle A.2: *Edges eines Album Node* (nach [Fac17e])

Edge	Type	Description	Note
picture	Picture	The cover photo of this album.	
photos	list<Photo>	Photos contained in this album.	
sharedposts	list<Post>	Stream posts that are shares of this album.	
likes	list<Profile>	People who like this.	
reactions	list<Profile>	People who have reacted to this.	
comments	Comment[]	Comments made on this.	

## A.2 Comment Node

Tabelle A.3: *Fields* eines *Comment Node* (nach [Fac17e])

Field	Type	Description	Note
attachment	StoryAttachment	Link or photo attached to the comment	
can_comment	bool	Whether the viewer can reply to this comment	special
can_hide	boolean	Whether the viewer can hide this comment. Only visible to a page admin	restricted
can_like	boolean	Whether the viewer can like this comment	special
can_remove	bool	Whether the viewer can remove this comment	special
can_reply_privately	boolean	Whether the viewer can send a private reply to this comment (Page viewers only)	special
comment_count	int32	Number of replies to this comment	
created_time	datetime	The time this comment was made	
from	User	The person that made this comment	
id	string	The comment ID	
like_count	int32	Number of times this comment was liked	
message	string	The comment text	
message_tags	object[]	An array of Profiles tagged in message.	
object	Object	For comments on a photo or video, this is that object. Otherwise, this is empty.	
parent	Comment	For comment replies, this is the comment that this is a reply to.	
private_reply_conversation	Conversation	For comments with private replies, gets conversation between the Page and author of the comment (Page viewers only)	restricted
user_likes	bool	Whether the viewer has liked this comment.	

Tabelle A.4: *Edges eines Comment Node* (nach [Fac17e])

Edge	Type	Description	Note
comments	Comment[]	Comments that reply to this comment.	
likes	list<User Page>	People who like this comment.	
reactions	list<Profile>	People who have reacted to this post.	
private_replies	n/a	Used to send private message reply to this comment (Page viewers only).	restricted

## A.3 Event Node

Tabelle A.5: *Fields eines Event Node* (nach [Fac17e])

Field	Type	Description	Note
attending_count	int32	Number of people attending the event	
can_guests_invite	bool	Can guests invite friends	
can_viewer_post	bool	Whether a user has permission to post on the event timeline	
category	string	The category of the event	
cover	CoverPhoto	Cover picture	
declined_count	int32	Number of people who declined the event	
description	string	Long-form description	
end_time	string	End time, if one has been set	
guest_list_enabled	bool	Can see guest list	
id	numeric string	The event ID	
interested_count	int32	Number of people interested in the event	
is_canceled	bool	Whether or not the event has been marked as canceled	
is_draft	bool	Whether the event is in draft mode or published	
is_page_owned	bool	Whether the event is created by page or not	
is_viewer_admin	bool	Whether the viewer is admin or not	
maybe_count	int32	Number of people who maybe going to the event	
name	string	Event name	
noreply_count	int32	Number of people who did not reply to the event	
owner	Profile	The profile that created the event	
parent_group	Group	The group the event belongs to	
place	Place	Event Place information	
start_time	string	Start time	
ticket_uri	string	The link users can visit to buy a ticket to this event	special
ticketing_privacy_uri	string	URI to seller's privacy policy for ticket purchases	special

Field	Type	Description	Note
ticketing_terms_uri	string	URI to seller's terms of service for ticket purchases	special
timezone	enum	Timezone	
type	enum{private, public, group, community}	The type of the event	
updated_time	datetime	Last update time (ISO 8601 formatted)	

Tabelle A.6: *Edges eines Event Node* (nach [Fac17e])

Edge	Type	Description	Note
admins	list<Profile>	Users that are admins of the event	
attending	list<Profile>	Users attending the event	
comments	Comment[]	Comments on the event	
declined	list<Profile>	Users who declined the event invitation	
interested	list<Profile>	Users interested in the event	
live_videos	list<LiveVideo>	Live videos published to the event	special
maybe	list<Profile>	Users who replied maybe to the event invitation	
noreply	list<Profile>	Users who have not replied to the event invitation	
photos	list<Photo>	Photos published to the event	
picture	Picture	Profile picture	
roles	list<Profile>	List of profiles having roles on the event	
feed	Post[]	The feed of posts (including status updates) and links published to this event's wall.	

## A.4 Page Node

Tabelle A.7: *Fields* eines *Page Node* (nach [Fac17e])

Field	Type	Description	Note
about	string	Information about the Page	
access_token	string	The access token you can use to act as the Page. Only visible to Page Admins.	restricted
ad_campaign	AdCampaign	The Page's currently running promotion campaign	restricted
affiliation	string	Affiliation of this person. Applicable to Pages representing people	special
app_id	id	App ID for app-owned Pages and app Pages	special
app_links	AppLinks	AppLinks data associated with the Page's URL	special
artists_we_like	string	Artists the band likes. Applicable to Bands	special
attire	string	Dress code of the business. Applicable to Restaurants or Nightlife. Can be one of Casual, Dressy or Unspecified	special
awards	string	The awards information of the film. Applicable to Films	special
band_interests	string	Band interests. Applicable to Bands	special
band_members	string	Members of the band. Applicable to Bands	special
best_page	Page	The best available Page on Facebook for the concept represented by this Page.	
bio	string	Biography of the band. Applicable to Bands	special
birthday	string	Birthday of this person. Applicable to Pages representing people	special
booking_agent	string	Booking agent of the band. Applicable to Bands	special
built	string	Year vehicle was built. Applicable to Vehicles	special
business	n/a	The Business associated with this Page. Visible only with a page access token or a user access token that has admin rights on the page	restricted

Field	Type	Description	Note
can_checkin	bool	Whether this page has checkin functionality enabled	
can_post	bool	Whether the current session user can post on this Page	
category	string	The Page's category. e.g. Product/Service, Computers/Technology	
category_list	list<PageCategory>	The Page's sub-categories	
checkins	unsigned int32	Number of checkins at a place represented by a Page	
company_overview	string	The company overview. Applicable to Companies	special
contact_address	MailingAddress	The mailing or contact address for this page. This field will be blank if the contact address is the same as the physical address	
context	OpenGraphContext	Social context for this Page	
country_page_likes	unsigned int32	If this is a Page in a Global Pages hierarchy, the number of people who are being directed to this Page.	special
cover	CoverPhoto	Information about the page's cover photo	
culinary_team	string	Culinary team of the business. Applicable to Restaurants or Nightlife	special
current_location	string	Current location of the Page	
description	string	The description of the Page	
description_html	string	The description of the Page in raw HTML	special
directed_by	string	The director of the film. Applicable to Films	special
display_subtext	string	Subtext about the Page being viewed	
displayed_message_response_time	string	Page estimated message response time displayed to user	
emails	list<string>	The emails listed in the About section of a Page	
engagement	Engagement	The social sentence and like count information for this Page. This is the same info used for the like button	special



Field	Type	Description	Note
fan_count	unsigned int32	The number of users who like the Page. For Global Pages this is the count for all Pages across the brand.	
featured_video	Video	Video featured by the Page	
features	string	Features of the vehicle. Applicable to Vehicles	special
food_styles	list<string>	The restaurant's food styles. Applicable to Restaurants	special
founded	string	When the company was founded. Applicable to Pages in the Company category	special
general_info	string	General information provided by the Page	
general_manager	string	General manager of the business. Applicable to Restaurants or Nightlife	special
genre	string	The genre of the film. Applicable to Films	special
global_brand_page_name	string	The name of the Page with country codes appended for Global Pages. Only visible to the Page admin	restricted
global_brand_root_id	numeric string	This brand's global Root ID	special
has_added_app	bool	Indicates whether this Page has added the app making the query in a Page tab	special
hometown	string	Hometown of the band. Applicable to Bands	special
hours	map<string, string>	Indicates a single range of opening hours for a day. Each day can have 2 different hours ranges.	special
id	numeric string	Page ID. No access token is required to access this field	
impressum	string	Legal information about the Page publishers	
influences	string	Influences on the band. Applicable to Bands	special
instant_articles_review_status	enum	Indicates the current Instant Articles review status for this page	restricted
is_always_open	bool	Indicates whether this location is always open	special
is_community_page	bool	Indicates whether the Page is a community Page	

Field	Type	Description	Note
is_eligible_for_branded_content	bool	Indicates whether the page is eligible for the branded content tool	special
is_permanently_closed	bool	Whether the business corresponding to this Page is permanently closed	special
is_published	bool	Indicates whether the Page is published and visible to non-admins	special
is_unclaimed	bool	Indicates whether the Page is unclaimed	special
is_verified	bool	Deprecated, use „verification_status“. Pages with a large number of followers can be manually verified by Facebook as having an authentic identity. This field indicates whether the page is verified by this process.	deprecated
is_webhooks_subscribed	bool	Indicates whether the application is subscribed for real time updates from this page	special
keywords	null	Deprecated. Returns null	deprecated
leadgen_form_preview_details	LeadGenForm↔PreviewDetails	The details needed to generate an accurate preview of a lead gen form	special
leadgen_has_crm_integration	bool	Indicates whether this page hasApp subscribes page leads realtime update	special
leadgen_tos_acceptance_time	datetime	Indicates the time when the TOS for running LeadGen Ads on the page was accepted	special
leadgen_tos_accepted	bool	Indicates whether a user has accepted the TOS for running LeadGen Ads on the Page	special
leadgen_tos_accepting_user	User	Indicates the user who accepted the TOS for running LeadGen Ads on the page	special
link	string	The Page's Facebook URL	
location	Location	The location of this place. Applicable to all Places	special
members	string	Members of this org. Applicable to Pages representing Team Orgs	special

Field	Type	Description	Note
merchant_id	string	The instant workflow merchant id associated with the Page	special
merchant_review_status	enum	Review status of the Page against FB commerce policies, this status decides whether the Page can use component flow	special
mission	string	The company mission. Applicable to Companies	special
mpg	string	MPG of the vehicle. Applicable to Vehicles	special
name	string	The name of the Page	restricted
name_with_location_descriptor	string	The name of the Page with its location and/or global brand descriptor. Only visible to a page admin. Non-page admins will get the same value as name.	
network	string	The TV network for the TV show. Applicable to TV Shows	special
new_like_count	unsigned int32	The number of people who have liked the Page, since the last login. Only visible to a page admin	restricted
offer_eligible	bool	Offer eligibility status. Only visible to a page admin	restricted
overall_star_rating	float	Overall page rating based on rating survey from users on a scale of 1-5. This value is normalized and is not guaranteed to be a strict average of user ratings.	
parent_page	Page	Parent Page for this Page	special
parking	PageParking	Parking information. Applicable to Businesses and Places	special
payment_options	PagePayment↔Options	Payment options accepted by the business. Applicable to Restaurants or Nightlife	special
personal_info	string	Personal information. Applicable to Pages representing People	special
personal_interests	string	Personal interests. Applicable to Pages representing People	special

Field	Type	Description	Note
pharma_safety_info	string	Pharmacy safety information. Applicable to Pharmaceutical companies	special
phone	string	Phone number provided by a Page	
place_type	enum	For places, the category of the place	special
plot_outline	string	The plot outline of the film. Applicable to Films	special
preferred_audience	Targeting	Group of tags describing the preferred audience of ads created for the Page	restricted
press_contact	string	Press contact information of the band. Applicable to Bands	special
price_range	string	Price range of the business. Applicable to Restaurants or Nightlife.	special
produced_by	string	The producer of the film. Applicable to Films	special
products	string	The products of this company. Applicable to Companies	special
promotion_eligible	bool	Reason why a post isn't eligible for boosting. Only visible to Page Admins	restricted
promotion_ineligible_reason	string	Reason, for which boosted posts are not eligible. Only visible to a page admin	restricted
public_transit	string	Public transit to the business. Applicable to Restaurants or Nightlife	special
publisher_space	PublisherSpace	Publisher Space for the page.	special
rating_count	unsigned int32	Number of ratings for the page.	
recipient	numeric string	Messenger page scope id associated with page and a user using account_linking_token	special
record_label	string	Record label of the band. Applicable to Bands	special
release_date	string	The film's release date. Applicable to Films	special
restaurant_services	PageRestaurant↔Services	Services the restaurant provides. Applicable to Restaurants	special
restaurant_specialties	PageRestaurant↔Specialties	The restaurant's specialties. Applicable to Restaurants	special

Field	Type	Description	Note
schedule	string	The air schedule of the TV show. Applicable to TV Shows	special
screenplay_by	string	The screenwriter of the film. Applicable to Films	special
season	string	The season information of the TV Show. Applicable to TV Shows	special
single_line_address	string	The page address, if any, in a simple single line format.	
starring	string	The cast of the film. Applicable to Films	special
start_info	PageStartInfo	Information about when the entity represented by the Page was started	
store_location_descriptor	string	Location Page's store location descriptor	special
store_number	unsigned int32	Unique store number for this location Page	special
studio	string	The studio for the film production. Applicable to Films	special
supports_instant_articles	bool	Indicates whether this Page supports Instant Articles	special
talking_about_count	unsigned int32	The number of people talking about this Page	
unread_message_count	unsigned int32	Unread message count for the Page. Only visible to a page admin	restricted
unread_notif_count	unsigned int32	Number of unread notifications. Only visible to a page admin	restricted
unseen_message_count	unsigned int32	Unseen message count for the Page. Only visible to a page admin	restricted
username	string	The alias of the Page. For example, for <a href="https://www.facebook.com/platform">www.facebook.com/platform</a> the username is 'platform'	
verification_status	string	Showing whether this Page is verified and in what color e.g. blue verified, gray verified or not verified	
voip_info	VoipInfo	Voip info	special
website	string	The URL of the Page's website	

Field	Type	Description	Note
were_here_count	unsigned int32	The number of visits to this Page's location. If the Page setting Show map, check-ins and star ratings on the Page (under Page Settings > Page Info > Address) is disabled, then this value will also be disabled	special
written_by	string	The writer of the TV show. Applicable to TV Shows	special

Tabelle A.8: Edges eines Page Node (nach [Fac17e])

Edge	Type	Description	Note
admin_notes	list <PageAdminNote>	Notes of this page	restricted
albums	list<Album>	Photo albums for this Page	
bc_sponsored_posts	list <EntWithSponsor>	A list of posts where the page is tagged by a branded content sponsor.	special
blocked	list<Profile>	User or Page Profiles blocked from this Page	restricted
business_activities	list<Business↔ ActivityLogEvent>	The business activities related to this page	restricted
call_to_actions	list<Page↔ CallToAction>	The call-to-action created by this Page	special
canvas_elements	list<Canvas↔ BodyElement>	The canvas elements associated with this page	special
conversations	list<UnifiedThread>	This Page's conversations	deprecated
crosspost_pending_approval_pages	list<Page>	Pages that have not given approval for crossposting	restricted
crosspost_whitelisted_pages	list<Page>	Pages whitelisted for crossposting	restricted
curated_collections	list<Curated↔ Collections>	The curated collections created by this page	special
events	list<Event>	Events for this Page	
featured_videos_collection	list<Video>	Featured Videos for this Page.	restricted
feed	Post[]	The feed of posts (including status updates) and links published by this page, or by others on this page.	redundant
global_brand_children	list<Page>	Children Pages of a Global Page's root Page. Only root Page can return children Pages.	restricted

Edge	Type	Description	Note
insights	list<InsightsResult>	This Page's Insights data. Global brand default Page will return its own insights data, while root Page will return integrated insights of the whole hierarchy.	restricted
instagram_accounts	list<InstagramUser>	Linked Instagram accounts for this Page	restricted
instant_articles	list<InstantArticle>	Instant articles associated with this Page	restricted
instant_articles_insights	list<Insights↔QueryResult>	Instant Article Aggregated Insights	restricted
labels	list<PageLabel>	Labels of the Page	restricted
leadgen_conditional_questions_group	list<LeadGen↔Conditional↔QuestionsGroup>	Creates conditional question.	restricted
leadgen_context_cards	list<LeadGen↔ContextCard>	A library of leadgen context cards that can be reused across forms	restricted
leadgen_legal_content	list<LeadGen↔LegalContent>	A library of leadgen legal content that can be reused across forms.	restricted
leadgen_whitelisted_users	list<User>	List of all whitelisted users of a page.	restricted
likes	list<Page>	The Pages that like this Page.	
live_videos	list<LiveVideo>	Live videos from this page	restricted
locations	Page[]	The location Pages that are children of this Page	special
media_fingerprints	list<Media↔Fingerprint>	Media fingerprints from this page	special
messenger_profile	n/a	messenger profile	restricted
milestones	list<LifeEvent>	This Page's milestones	restricted
nativeoffers	list<NativeOffer>	The native offers created by this Page	restricted
notes	n/a	A Facebook Note created by a user or page.	deprecated
notifications	Notification[]	page/notifications API	restricted
offers_v3	n/a	offers v3	special
page_backed_instagram_accounts	list<InstagramUser>	Linked Page Backed Instagram accounts for this Page	restricted
photos	list<Photo>	This Page's photos	
picture	Picture	This Page's profile picture	
place_topics	list<PlaceTopic>	The place topic associated with the categories of the page	special

Edge	Type	Description	Note
posts	Post[]	The feed of posts (including status updates) and links published by this page, or by others on this page.	
product_catalogs	list <ProductCatalog>	Product catalogs owned by this page	special
promotable_posts	Post[]	The feed of posts (including status updates) and links published by this page, or by others on this page.	restricted
questions	n/a	Graph API Reference Question /question	deprecated
ratings	list<Open↔ GraphRating>	Open Graph ratings given to this Page	restricted
roles	list<User>	The Page's Admins	restricted
saved_filters	list <PageSavedFilter>	Filters saved in Page Publishing Tools	restricted
saved_message_responses	list <SavedMessage↔ Response>	All message responses that are saved by a page	restricted
screennames	list<ScreenName>	External accounts. Applicable to Pages representing people	special
seasons	list<VideoList>	Seasons for this show page	special
settings	list<PageSettings>	Controllable settings for this page	restricted
show_playlists	list<VideoList>	Playlists for this show page	special
statuses	n/a	The feed of status messages published by this page.	deprecated
subscribed_apps	list<Application>	Applications that have real time update subscriptions for this Page. Note that we will only return information about the current app	restricted
tabs	list<Tab>	This Page's tabs and the apps in them	restricted
tagged	list<Post>	The photos, videos, and posts in which the Page has been tagged. A derivative of /feeds	
thread_settings	list<ThreadSetting>	Messenger platform configs for the page	special
threads	n/a	A messages thread in Facebook Messenger	restricted
tours	list<EventTour>	tours	special
video_broadcasts	list<LiveVideo>	Video broadcasts from this page	restricted



Edge	Type	Description	Note
video_copyright_rules	list<Video↔CopyrightRule>	Video copyright rules from this page	restricted
video_copyrights	list<VideoCopyright>	Video copyrights from this page	restricted
video_lists	list<VideoList>	Video Playlists for this Page	
videos	list<Video>	Videos for this Page	
videos_you_can_use	list<Page↔VideosYouCanUse>	Provides a list of videos the Page can use for crossposting. Owners of these videos have granted permission for the Page to crosspost them.	restricted
workflows	n/a	Native component workflows available for this Page. Workflows allow people to perform actions like requesting appointments, buying tickets, and ordering from restaurants on Facebook.	special

## A.5 Photo Node

Tabelle A.9: *Fields* eines *Photo Node* (nach [Fac17e])

Field	Type	Description	Note
album	Album	The album this photo is in	
backdated_time	datetime	A user-specified time for when this object was created	special
backdated_time_granularity	enum	How accurate the backdated time is	special
can_backdate	bool	Indicates whether the viewer can backdate the photo	special
can_delete	bool	Indicates whether the viewer can delete the photo	special
can_tag	bool	Indicates whether the viewer can tag the photo	special
created_time	datetime	The time this photo was published	
event	Event	If this object has a place, the event associated with the place	
from	User Page	The profile (user or page) that uploaded this photo	
height	unsigned int32	The height of this photo in pixels	
icon	string	The icon that Facebook displays when photos are published to News Feed	special
id	numeric string	The photo ID	
images	list<Platform↔ImageSource>	The different stored representations of the photo. Can vary in number based upon the size of the original photo.	
link	string	A link to the photo on Facebook	
name	string	The user-provided caption given to this photo. Corresponds to caption when creating photos	

Field	Type	Description	Note
name_tags	EntityAtText↔ Range[]	An array containing an array of objects mentioned in the name field which contain the id, name, and type of each object as well as the offset and length which can be used to match it up with its corresponding string in the name field	
page_story_id	string	ID of the page story this corresponds to. May not be on all photos. Applies only to published photos	special
picture	string	Link to the 100px wide representation of this photo	
place	Place	Location associated with the photo, if any	
position	unsigned int32	Deprecated. Returns 0	deprecated
source	string	Deprecated. Use images instead	deprecated
target	Profile	The target this photo is published to	
updated_time	datetime	The last time the photo was updated	
webp_images	list<Platform↔ ImageSource>	The different stored representations of the photo in webp format. Can vary in number based upon the size of the original photo.	
width	unsigned int32	The width of this photo in pixels	

Tabelle A.10: Edges eines Photo Node (nach [Fac17e])

Edge	Type	Description	Note
comments	Comment[]	Comments on an object	
insights	list<InsightsResult>	Insights data	restricted
likes	list<Profile>	People who like this	
reactions	list<Profile>	People who reacted on this	
sharedposts	list<Post>	The posts in which this photo is shared	
sponsor_tags	list<Page>	Sponsor pages tagged in the photo.	special
tags	list<Profile>	The Users tagged in the photo	

## A.6 Picture Node

Tabelle A.11: *Fields* eines *Picture Node* (nach [Fac17e])

Field	Type	Description	Note
bottom	unsigned int32	Deprecated	deprecated
height	unsigned int32	Picture height in pixels. Only returned when specified as a modifier	
is_silhouette	bool	True if the profile picture is the default 'silhouette' picture	
left	unsigned int32	Deprecated	deprecated
right	unsigned int32	Deprecated	deprecated
top	unsigned int32	Deprecated	deprecated
url	string	URL of the profile picture	
width	unsigned int32	Picture width in pixels. Only returned when specified as a modifier	

Tabelle A.12: *Edges* eines *Picture Node* (nach [Fac17e])

Der *Picture Node* besitzt keine weiteren *Edges*.

## A.7 Post Node

Tabelle A.13: *Fields* eines *Post Node* (nach [Fac17e])

Field	Type	Description	Note
admin_creator	object[]	ID of admin, app or business that created the post. Applies to pages only	special
application	App	Information about the app this post was published by.	special
call_to_action	object	The call to action type used in any Page posts for mobile app engagement ads.	special
caption	string	Link caption in post that appears below name. The caption must be an actual URLs and should accurately reflect the URL and associated advertiser or business someone visits when they click on it.	special
created_time	datetime	The time the post was initially published. For a post about a life event, this will be the date and time of the life event	
description	string	A description of a link in the post (appears beneath the caption).	
feed_targeting	object	Object that controls news feed targeting for this post. Anyone in these groups will be more likely to see this post, others will be less likely, but may still see it anyway. Any of the targeting fields shown here can be used, none are required (applies to Pages only).	special
from	Profile	Information about the profile that posted the message.	
icon	string	A link to an icon representing the type of this post.	special
id	string	The post ID	

Field	Type	Description	Note
instagram_eligibility	string	Whether the post can be promoted on Instagram. It returns the enum „eligible“ if it can be promoted. Otherwise it returns an enum for why it cannot be promoted	special
is_hidden	boolean	If this post is marked as hidden (Applies to Pages only).	special
is_instagram_eligible	string	Whether this post can be promoted in Instagram	special
is_published	boolean	Indicates whether a scheduled post was published (applies to scheduled Page Post only, for users post and instantly published posts this value is always true). Note that this value is always false for page posts created as part of the Ad Creation process.	special
link	string	The link attached to this post.	
message	string	The status message in the post.	
message_tags	object	Profiles tagged in message. This is an object with a unique key for each tag in the message	
name	string	The name of the link.	
object_id	string	The ID of any uploaded photo or video attached to the post.	
parent_id	string	The ID of a parent post for this post, if it exists. For example, if this story is a 'Your Page was mentioned in a post' story, the parent_id will be the original post where the mention happened	
permalink_url	string	URL to the permalink page of the post.	
picture	string	The picture scraped from any link included with the post.	
place	Place	Any location information attached to the post.	
privacy	object	The privacy settings of the post.	

Field	Type	Description	Note
promotable_id	string	ID of post to use for promotion for stories that cannot be promoted directly	special
promotion_status	string	Status of the promotion, if the post was promoted. Requires Page admin privileges.	restricted
properties	object[]	A list of properties for any attached video, for example, the length of the video.	special
shares	object	The shares count of this post. The share count may include deleted posts and posts you cannot see for privacy reasons.	
source	string	A URL to any Flash movie or video file attached to the post.	special
status_type	enum {mobile_status_update, created_note, added_photos, added_video, shared_story, created_group, created_event, wall_post, app_created_story, published_story, tagged_in_photo, approved_friend}	Description of the type of a status update.	
story	string	Text from stories not intentionally generated by users, such as those generated when two people become friends, or when someone else posts on the person's wall.	
story_tags	array	Deprecated field, same as message_tags.	deprecated

Field	Type	Description	Note
targeting	object	Object that limits the audience for this content. Only audiences in the specified demographics can view this content. The demographics are additive. Each additional value adds its audience to the cumulative targeted audience. These values do not override any Page-level demographic restrictions that may be in place.	special
to	Profile[]	Profiles mentioned or targeted in this post.	
type	enum{link, status, photo, video, offer}	A string indicating the object type of this post.	
updated_time	datetime	The time when the post was created, last edited or the time of the last comment that was left on the post. For a post about a life event, this will be the date and time of the life event	
with_tags	JSON object with a data field that contains a list of Profile objects.	Profiles tagged as being 'with' the publisher of the post.	

Tabelle A.14: Edges eines Post Node (nach [Fac17e])

Edge	Type	Description	Note
attachments	list<Attachment>	All media attachments associated with this post.	
comments	Comment[]	Comments on this post.	
insights	list<InsightsResult>	Insights for this post (only for Pages).	restricted
likes	list<Profile>	People who like this post.	
reactions	list<Profile>	People who have reacted to this post.	
sharedposts	list<Post>	Shares of this post.	



## A.8 User Node

Tabelle A.15: *Fields* eines *User Node* (nach [Fac17e])

Field	Type	Description	Note
about	string	Equivalent to the bio field	restricted
admin_notes	list <PageAdminNote>	Notes added by viewing page on this person	restricted
age_range	AgeRange	The age segment for this person expressed as a minimum and maximum age. For example, more than 18, less than 21.	restricted
birthday	string	The person's birthday. This is a fixed format string, like MM/DD/YYYY. However, people can control who can see the year they were born separately from the month and day so this string can be only the year (YYYY) or the month + day (MM/DD)	restricted
context	UserContext	Social context for this person	restricted
cover	CoverPhoto	The person's cover photo	
currency	Currency	The person's local currency information	restricted
devices	list<UserDevice>	The list of devices the person is using. This will return only iOS and Android devices	restricted
education	list<Education↔Experience>	The person's education	restricted
email	string	The person's primary email address listed on their profile. This field will not be returned if no valid email address is available	restricted
employee_number	string	The person's employee number, as set by the company via SCIM API	restricted
favorite_athletes	list<Experience>	Athletes the person likes	restricted
favorite_teams	list<Experience>	Sports teams the person likes	restricted
first_name	string	The person's first name	

Field	Type	Description	Note
gender	string	The gender selected by this person, male or female. If the gender is set to a custom value, this value will be based off of the preferred pronoun; it will be omitted if the preferred preferred pronoun is neutral	restricted
hometown	Page	The person's hometown	restricted
id	numeric string	The id of this person's user account. This ID is unique to each app and cannot be used across different apps. Our upgrade guide provides more information about app-specific IDs	
inspirational_people	list<Experience>	The person's inspirational people	restricted
install_type	enum	Install type	restricted
installed	bool	Is the app making the request installed?	restricted
interested_in	list<string>	Genders the person is interested in	restricted
is_shared_login	bool	Is this a shared login (e.g. a gray user)	restricted
is_verified	bool	People with large numbers of followers can have the authenticity of their identity manually verified by Facebook. This field indicates whether the person's profile is verified in this way. This is distinct from the verified field	restricted
labels	list<PageLabel>	Labels applied by viewing page on this person	restricted
languages	list<Experience>	Facebook Pages representing the languages this person knows	restricted
last_name	string	The person's last name	
link	string	A link to the person's Timeline	restricted
local_news_megaphone_dismiss_status	bool	Display megaphone for local news bookmark	restricted
locale	string	The person's locale	restricted

Field	Type	Description	Note
location	Page	The person's current location as entered by them on their profile. This field is not related to check-ins	restricted
meeting_for	list<string>	What the person is interested in meeting for	restricted
middle_name	string	The person's middle name	
name	string	The person's full name	
name_format	string	The person's name formatted to correctly handle Chinese, Japanese, or Korean ordering	restricted
payment_pricepoints	PaymentPricepoints	The person's payment pricepoints	restricted
political	string	The person's political views	restricted
public_key	string	The person's PGP public key	restricted
quotes	string	The person's favorite quotes	restricted
relationship_status	string	The person's relationship status	restricted
religion	string	The person's religion	restricted
security_settings	SecuritySettings	Security settings	restricted
shared_login_upgrade_required_by	datetime	The time that the shared login needs to be upgraded to Business Manager by	restricted
significant_other	User	The person's significant other	restricted
sports	list<Experience>	Sports played by the person	restricted
test_group	unsigned int32	Platform test group	restricted
third_party_id	string	A string containing an anonymous, but unique identifier for the person. You can use this identifier with third parties	restricted
timezone	float (min: -24) (max: 24)	The person's current timezone offset from UTC	restricted
token_for_business	string	A token that is the same across a business's apps. Access to this token requires that the person be logged into your app or have a role on your app. This token will change if the business owning the app changes	restricted

Field	Type	Description	Note
updated_time	datetime	Updated time	restricted
verified	bool	Indicates whether the account has been verified. This is distinct from the is_verified field.	restricted
video_upload_limits	VideoUploadLimits	Video upload limits	restricted
viewer_can_send_gift	bool	Can the viewer send a gift to this person?	restricted
website	string	The person's website	restricted
work	list<WorkExperience>	Details of a person's work experience	restricted

Tabelle A.16: Edges eines User Node (nach [Fac17e])

Edge	Type	Description	Note
accounts	list<Page>	Facebook Pages this person administers/is an admin for	restricted
achievements	list<OpenGraphAction:games.achieves>	Achievements made in Facebook games	restricted
ad_studies	list<AdStudy>	Ad studies that this person can view	restricted
adaccounts	list<AdAccount>	The advertising accounts to which this person has access	restricted
adcontracts	list<AdContract>	The person's ad contracts	restricted
adnetworkanalytics	list<InsightsQueryResult>	Insights data for the person's Audience Network apps	restricted
albums	list<Album>	The photo albums this person has created	restricted
apprequest↔ formerrecipients	list<AppRequestFormerRecipient>	App requests	restricted
apprequests	list<AppRequest>	This person's pending requests from an app	restricted
books	list<Page>	The books listed on this person's profile	restricted
business_activities	list<BusinessActivityLogEvent>	The business activities related to this user	restricted
business_users	list<BusinessUser>	Business users corresponding to the user	restricted
businesses	list<Business>	Businesses associated with the user	restricted
checkins	n/a	The checkins this person has made.	deprecated

Edge	Type	Description	Note
conversations	list<UnifiedThread>	Facebook Messenger conversation	restricted
curated_collections	list<CuratedCollection>	The curated collections created by this user	restricted
domains	list<Domain>	The domains the user admin	restricted
events	list<Event>	Events for this person. By default this does not include events the person has declined or not replied to	restricted
family	list<User>	This person's family relationships.	restricted
favorite_requests	list<FavoriteRequest>	Developers' favorite requests to the Graph API	restricted
feed	Post[]	The feed of posts (including status updates) and links published by this person.	restricted
friendlists	list<FriendList>	The person's custom friend lists	restricted
friendrequests	n/a	A person's pending friend requests.	deprecated
friends	list<User>	A person's friends.	restricted
games	list<Page>	Games this person likes	restricted
groups	list<Group>	The Facebook Groups that the person is an admin of	restricted
home	Post[]	A person's Facebook home-page feed.	deprecated
ids_for_apps	list<UserIDForApp>	Businesses can claim ownership of multiple apps using Business Manager. This edge returns the list of IDs that this user has in any of those other apps	restricted
ids_for_business	list<UserIDForApp>	Businesses can claim ownership of multiple apps using Business Manager. This edge returns the list of IDs that this user has in any of those other apps	restricted
ids_for_pages	list<UserIDForPage>	Businesses can claim ownership of apps and pages using Business Manager. This edge returns the list of IDs that this user has in any of the pages owned by this business.	restricted
inbox	Thread[]	A person's Facebook Messages inbox.	deprecated

Edge	Type	Description	Note
invitable_friends	list<User↔ InvitableFriend>	A list of friends that can be invited to install a Facebook Canvas app	restricted
likes	list<Page>	All the Pages this person has liked	restricted
live_videos	list<LiveVideo>	Live videos from this person	restricted
locations	Post[] Photo[]	A feed of posts and photos that include location information and in which this person has been tagged. This is useful for constructing a chronology of places that the person has visited.	deprecated
movies	list<Page>	Movies this person likes	restricted
music	list<Page>	Music this person likes	restricted
mutualfriends	n/a	The list of mutual friends between two people.	deprecated
notifications	n/a	The unread Facebook notifications that a person has.	restricted
objects	list<OpenGraph↔ Object:generic>	Objects	restricted
outbox	Thread[]	A person's Facebook Messages outbox.	restricted
permissions	list<permission>	The permissions that the person has granted this app	restricted
personal_ad_account	list<AdAccount>	The advertising accounts to which this person has personal access	restricted
photos	list<Photo>	Photos the person is tagged in or has uploaded	restricted
picture	Picture	The person's profile picture	
promotable_domains	list<Domain>	All the domains user can promote	restricted
promotable_events	list<Event>	All the events which user can promote.	restricted
questions		The questions that a person has created.	restricted
request_history	list <RequestHistory>	Developers' Graph API request history	restricted
scores	n/a	The scores this person has received from Facebook Games that they've played.	deprecated
session_keys	list<Platform↔ SessionKey>	Any session keys that the app knows the user by	restricted
stream_filters	list<StreamFilter>	A list of filters that can be applied to the News Feed edge	restricted

Edge	Type	Description	Note
subscribedto	n/a	The profile that this person is following.	deprecated
subscribers	n/a	The profiles that are following this person.	deprecated
taggable_friends	list<User↔ TaggableFriend>	Friends that can be tagged in content published via the Graph API	restricted
tagged_places	list<PlaceTag>	List of tagged places for this person. It can include tags on videos, posts, statuses or links	restricted
television	list<Page>	TV shows this person likes	restricted
threads	list<UnifiedThread>	A message conversation thread	restricted
video_broadcasts	list<LiveVideo>	Video broadcasts from this person	restricted
videos	list<Video>	Videos the person is tagged in or uploaded	restricted

## A.9 Video Node

Tabelle A.17: *Fields* eines *Video Node* (nach [Fac17e])

Field	Type	Description	Note
ad_breaks	list<integer>	Time offsets of ad breaks in milliseconds. Ad breaks are short ads that play within a video.	special
backdated_time	datetime	The time when the video post was created.	special
backdated_time_granularity	enum	Accuracy of the backdated time.	special
content_category	enum	The content category of this video.	
content_tags	list<numeric string>	Tags that describe the contents of the video.	
created_time	datetime	The time the video was initially published.	
custom_labels	list<string>	Labels used to describe the video. Unlike content tags, custom labels are not published and only appear in insights data.	restricted
description	string	The description of the video.	
embed_html	string	The HTML element that may be embedded in a Web page to play the video.	special
embeddable	bool	Whether the video is embeddable.	special
event	Event	If this object has a place, the event associated with the place	special
format	list<VideoFormat>	The different formats of the video.	
from	Profile	The profile that created the video.	
icon	string	The icon that Facebook displays when videos are published to the feed.	special
id	numeric string	The video ID.	
is_crosspost_video	bool	Whether the video is cross-posted from other page.	special
is_crossposting_eligible	bool	Specifies if the video is eligible for crossposting. Page access-token is required to query this field.	restricted



Field	Type	Description	Note
is_instagram_eligible	bool	Whether the video is eligible to be promoted on Instagram	special
length	float	Duration of this video.	
live_status	enum	The live status of the video	
monetization_status	enum	The monetization status of this video	special
permalink_url	string	URL to the permalink page of the video.	
picture	string	The URL for the thumbnail picture of the video.	special
place	Place	Location associated with the video, if any.	
privacy	Privacy	Privacy setting for the video.	
published	bool	Whether a post about this video is published. The post is not scheduled, draft, or ads_post.	special
scheduled_publish_time	datetime	The time that the video is scheduled to publish.	special
source	string	A URL to the raw, playable video file.	
status	VideoStatus	Object describing the status attributes of a video.	
title	string	The video title or caption.	
universal_video_id	string	The publishers asset management code for this video.	special
updated_time	datetime	The last time the video or its caption was updated.	

Tabelle A.18: *Edges eines Video Node* (nach [Fac17e])

Edge	Type	Description	Note
auto_generated_captions	list<VideoAutoGeneratedCaption>	Auto-generated Captions from automated speech recognition system.	restricted
captions	list<VideoCaption>	Captions for the video.	
comments	list<Comment>	Users comments for the video.	
crosspost_shared_pages	list<Page>	The pages that this video being crossposting shared to.	special
likes	list<Profile>	People who like this	
reactions	list<Profile>	People who reacted on this	
sharedposts	list<Post>	Posts that share this video.	
sponsor_tags	list<Page>	Sponsor pages tagged in the video.	special
tags	list<Profile>	Users tagged in the video.	
thumbnails	list<VideoThumbnail>	Thumbnails for the video.	special
video_insights	list<InsightsResult>	Total insights from all video posts associated with this video.	restricted

## A.10 VideoList Node

Tabelle A.19: *Fields eines VideoList Node* (nach [Fac17e])

Field	Type	Description	Note
creation_time	datetime	The time when the playlist was created	
description	string	Description of the playlist	
id	numeric string	Video playlist ID	
last_modified	datetime	The time when the contents of the playlist was last changed	
owner	User Page	Owner of the playlist	
thumbnail	string	Thumbnail of the playlist	
title	string	Title of the playlist	
videos_count	int32	Number of Videos in the playlist	

Tabelle A.20: *Edges eines VideoList Node* (nach [Fac17e])

Edge	Type	Description	Note
videos	list<Video>	Videos in the playlist	

## Anhang B: Feature-Kandidaten relevanter Facebook Nodes

### B.1 Album Node

Tabelle B.1: Feature-Kandidaten der *Fields* eines *Album Node* (nach [Fac17e])

Field	Type	Description	Note
cover_photo	Photo	The ID of the album's cover photo.	
created_time	datetime	The time the album was initially created.	
description	string	The description of the album.	considered
event	Event	The event associated with this album.	
from	User	The profile that created the album.	meta
id	string	The album ID.	meta
link	string	A link to this album on Facebook.	
location	string	The textual location of the album.	
name	string	The title of the album.	considered
place	Page	The place associated with this album.	
privacy	string	The privacy settings for the album.	
type	enum {app, cover, profile, mobile, wall, normal, album}	The type of the album.	
updated_time	datetime	The last time the album was updated.	

Tabelle B.2: Feature-Kandidaten der *Edges* eines *Album Node* (nach [Fac17e])

Edge	Type	Description	Note
picture	Picture	The cover photo of this album.	considered
photos	list<Photo>	Photos contained in this album.	considered
sharedposts	list<Post>	Stream posts that are shares of this album.	
likes	list<Profile>	People who like this.	
reactions	list<Profile>	People who have reacted to this.	
comments	Comment[]	Comments made on this.	partially considered

## B.2 Comment Node

Tabelle B.3: Feature-Kandidaten der *Fields* eines *Comment Node* (nach [Fac17e])

Field	Type	Description	Note
attachment	StoryAttachment	Link or photo attached to the comment	
comment_count	int32	Number of replies to this comment	meta
created_time	datetime	The time this comment was made	
from	User	The person that made this comment	meta
id	string	The comment ID	meta
like_count	int32	Number of times this comment was liked	meta
message	string	The comment text	considered
message_tags	object[]	An array of Profiles tagged in message.	
object	Object	For comments on a photo or video, this is that object. Otherwise, this is empty.	
parent	Comment	For comment replies, this the comment that this is a reply to.	
user_likes	bool	Whether the viewer has liked this comment.	

Tabelle B.4: Feature-Kandidaten der *Edges* eines *Comment Node* (nach [Fac17e])

Edge	Type	Description	Note
comments	Comment[]	Comments that reply to this comment.	
likes	list<User Page>	People who like this comment.	
reactions	list<Profile>	People who have reacted to this post.	

## B.3 Event Node

Tabelle B.5: Feature-Kandidaten der *Fields* eines *Event Node* (nach [Fac17e])

Field	Type	Description	Note
attending_count	int32	Number of people attending the event	meta
can_guests_invite	bool	Can guests invite friends	
can_viewer_post	bool	Whether a user has permission to post on the timeline	
category	string	The category of the event	considered
cover	CoverPhoto	Cover picture	
declined_count	int32	Number of people who declined the event	meta
description	string	Long-form description	considered
end_time	string	End time, if one has been set	
guest_list_enabled	bool	Can see guest list	
id	numeric string	The event ID	meta
interested_count	int32	Number of people interested in the event	meta
is_canceled	bool	Whether or not the event has been marked as canceled	
is_draft	bool	Whether the event is in draft mode or published	
is_page_owned	bool	Whether the event is created by page or not	
is_viewer_admin	bool	Whether the viewer is admin or not	
maybe_count	int32	Number of people who maybe going to the event	meta
name	string	Event name	considered
noreply_count	int32	Number of people who did not reply to the event	meta
owner	Profile	The profile that created the event	meta
parent_group	Group	The group the event belongs to	
place	Place	Event Place information	
start_time	string	Start time	
timezone	enum	Timezone	
type	enum{private, public, group, community}	The type of the event	
updated_time	datetime	Last update time (ISO 8601 formatted)	

Tabelle B.6: Feature-Kandidaten der *Edges* eines *Event Node* (nach [Fac17e])

Edge	Type	Description	Note
admins	list<Profile>	Users that are admins of the event	
attending	list<Profile>	Users attending the event	
comments	Comment[]	Comments on the event	partially considered
declined	list<Profile>	Users who declined the event invitation	
interested	list<Profile>	Users interested in the event	
maybe	list<Profile>	Users who replied maybe to the event invitation	
noreply	list<Profile>	Users who have not replied to the event invitation	
photos	list<Photo>	Photos published to the event	considered
picture	Picture	Profile picture	considered
roles	list<Profile>	List of profiles having roles on the event	
feed	Post[]	The feed of posts (including status updates) and links published to this event's wall.	partially considered

## B.4 Page Node

Tabelle B.7: Feature-Kandidaten der *Fields* eines *Page Node* (nach [Fac17e])

Field	Type	Description	Note
about	string	Information about the Page	considered
best_page	Page	The best available Page on Facebook for the concept represented by this Page.	meta
can_checkin	bool	Whether this page has checkin functionality enabled	
can_post	bool	Whether the current session user can post on this Page	
category	string	The Page's category. e.g. Product/Service, Computers/Technology	considered
category_list	list<PageCategory>	The Page's sub-categories	considered
checkins	unsigned int32	Number of checkins at a place represented by a Page	
contact_address	MailingAddress	The mailing or contact address for this page. This field will be blank if the contact address is the same as the physical address	
context	OpenGraphContext	Social context for this Page	
cover	CoverPhoto	Information about the page's cover photo	
current_location	string	Current location of the Page	
description	string	The description of the Page	considered
display_subtext	string	Subtext about the Page being viewed	
displayed_message_response_time	string	Page estimated message response time displayed to user	
emails	list<string>	The emails listed in the About section of a Page	
fan_count	unsigned int32	The number of users who like the Page. For Global Pages this is the count for all Pages across the brand.	meta
featured_video	Video	Video featured by the Page	considered
general_info	string	General information provided by the Page	considered
id	numeric string	Page ID. No access token is required to access this field	meta



Field	Type	Description	Note
impressum	string	Legal information about the Page publishers	considered
is_community_page	bool	Indicates whether the Page is a community Page	meta
link	string	The Page's Facebook URL	meta
name	string	The name of the Page	considered
overall_star_rating	float	Overall page rating based on rating survey from users on a scale of 1-5. This value is normalized and is not guaranteed to be a strict average of user ratings.	meta
phone	string	Phone number provided by a Page	
rating_count	unsigned int32	Number of ratings for the page.	meta
single_line_address	string	The page address, if any, in a simple single line format.	
start_info	PageStartInfo	Information about when the entity represented by the Page was started	
talking_about_count	unsigned int32	The number of people talking about this Page	meta
username	string	The alias of the Page. For example, for <a href="https://www.facebook.com/platform">www.facebook.com/platform</a> the username is 'platform'	considered
verification_status	string	Showing whether this Page is verified and in what color e.g. blue verified, gray verified or not verified	meta
website	string	The URL of the Page's website	meta

Tabelle B.8: Feature-Kandidaten der *Edges* eines *Page Node* (nach [Fac17e])

Edge	Type	Description	Note
albums	list<Album>	Photo albums for this Page	considered
events	list<Event>	Events for this Page	considered
feed	Post[]	The feed of posts (including status updates) and links published by this page, or by others on this page.	redundant
likes	list<Page>	The Pages that like this Page.	
photos	list<Photo>	This Page's photos	considered
picture	Picture	This Page's profile picture	considered
posts	Post[]	The feed of posts (including status updates) and links published by this page, or by others on this page.	considered
tagged	list<Post>	The photos, videos, and posts in which the Page has been tagged. A derivative of /feeds	
video_lists	list<VideoList>	Video Playlists for this Page	considered
videos	list<Video>	Videos for this Page	considered

## B.5 Photo Node

Tabelle B.9: Feature-Kandidaten der *Fields* eines *Photo Node* (nach [Fac17e])

Field	Type	Description	Note
album	Album	The album this photo is in	meta
created_time	datetime	The time this photo was published	
event	Event	If this object has a place, the event associated with the place	
from	User Page	The profile (user or page) that uploaded this photo	meta
height	unsigned int32	The height of this photo	meta
id	numeric string	The photo ID	meta
images	list<Platform↔ImageSource>	The different stored representations of the photo. Can vary in number based upon the size of original photo.	considered
link	string	Link to the photo on Facebook	meta
name	string	The user-provided caption given to this photo. Corresponds to caption when creating photos	considered
name_tags	EntityAtText↔Range[]	An array containing an array of objects mentioned in the name field which contain the id, name, and type of each object as well as the offset and length	
picture	string	Link to the 100px wide representation of this photo	
place	Place	Location associated with the photo, if any	
target	Profile	The target this photo is published to	
updated_time	datetime	The last time the photo was updated	
webp_images	list<Platform↔ImageSource>	The different stored representations of the photo in webp format. Can vary in number based upon the size of the original photo.	
width	unsigned int32	The width of this photo	meta

Tabelle B.10: Feature-Kandidaten der *Edges* eines *Photo Node* (nach [Fac17e])

Edge	Type	Description	Note
comments	Comment[]	Comments on an object	partially considered
likes	list<Profile>	People who like this	
reactions	list<Profile>	People who reacted on this	
sharedposts	list<Post>	The posts in which this photo is shared	
tags	list<Profile>	The Users tagged in the photo	

## B.6 Picture Node

Tabelle B.11: Feature-Kandidaten der *Fields* eines *Picture Node* (nach [Fac17e])

Field	Type	Description	Note
height	unsigned int32	Picture height in pixels. Only returned when specified as a modifier	meta
is_silhouette	bool	True if the profile picture is the default 'silhouette' picture	
url	string	URL of the profile picture	meta
width	unsigned int32	Picture width in pixels. Only returned when specified as a modifier	meta

Tabelle B.12: Feature-Kandidaten der *Edges* eines *Picture Node* (nach [Fac17e])

Der *Picture Node* besitzt keine weiteren *Edges*.

## B.7 Post Node

Tabelle B.13: Feature-Kandidaten der *Fields* eines *Post Node* (nach [Fac17e])

Field	Type	Description	Note
created_time	datetime	The time the post was initially published. For a post about a life event, this will be the date and time of the life event	
description	string	A description of a link in the post (appears beneath the caption).	considered
from	Profile	Information about the profile that posted the message.	meta
id	string	The post ID	meta
link	string	The link attached to this post.	meta
message	string	The status message in the post.	considered
message_tags	object	Profiles tagged in message. This is an object with a unique key for each tag in the message	
name	string	The name of the link.	considered
object_id	string	The ID of any uploaded photo or video attached to the post.	
parent_id	string	The ID of a parent post for this post, if it exists. For example, if this story is a 'Your Page was mentioned in a post' story, the parent_id will be the original post where the mention happened	
permalink_url	string	URL to the permalink page of the post.	meta
picture	string	The picture scraped from any link included with the post.	
place	Place	Any location information attached to the post.	
privacy	object	The privacy settings of the post.	

Field	Type	Description	Note
shares	object	The shares count of this post. The share count may include deleted posts and posts you cannot see for privacy reasons.	meta
status_type	enum {mobile_status_update, created_note, added_photos, added_video, shared_story, created_group, created_event, wall_post, app_created_story, published_story, tagged_in_photo, approved_friend}	Description of the type of a status update.	meta
story	string	Text from stories not intentionally generated by users, such as those generated when two people become friends, or when someone else posts on the person's wall.	
to	Profile[]	Profiles mentioned or targeted in this post.	
type	enum{link, status, photo, video, offer}	A string indicating the object type of this post.	meta
updated_time	datetime	The time when the post was created, last edited or the time of the last comment that was left on the post. For a post about a life event, this will be the date and time of the life event	
with_tags	JSON object with a data field that contains a list of Profile objects.	Profiles tagged as being 'with' the publisher of the post.	

Tabelle B.14: Feature-Kandidaten der *Edges* eines *Post Node* (nach [Fac17e])

Edge	Type	Description	Note
attachments	list<Attachment>	All media attachments associated with this post.	considered
comments	Comment[]	Comments on this post.	partially considered
likes	list<Profile>	People who like this post.	
reactions	list<Profile>	People who have reacted to this post.	
sharedposts	list<Post>	Shares of this post.	

## B.8 User Node

Tabelle B.15: Feature-Kandidaten der *Fields* eines *User Node* (nach [Fac17e])

Field	Type	Description	Note
cover	CoverPhoto	The person's cover photo	
first_name	string	The person's first name	
id	numeric string	The id of this person's user account. This ID is unique to each app and cannot be used across different apps. Our upgrade guide provides more information about app-specific IDs	
last_name	string	The person's last name	
middle_name	string	The person's middle name	
name	string	The person's full name	

Tabelle B.16: Feature-Kandidaten der *Edges* eines *User Node* (nach [Fac17e])

Edge	Type	Description	Note
picture	Picture	The person's profile picture	

## B.9 Video Node

Tabelle B.17: Feature-Kandidaten der *Fields* eines *Video Node* (nach [Fac17e])

Field	Type	Description	Note
content_category	enum	The content category of this video.	considered
content_tags	list<numeric string>	Tags that describe the contents of the video.	considered
created_time	datetime	The time the video was initially published.	
description	string	The description of the video.	considered
format	list<VideoFormat>	The different formats of the video.	meta
from	Profile	The profile that created the video.	meta
id	numeric string	The video ID.	meta
length	float	Duration of this video.	meta
live_status	enum	The live status of the video	
permalink_url	string	URL to the permalink page of the video.	meta
place	Place	Location associated with the video, if any.	
privacy	Privacy	Privacy setting for the video.	
source	string	A URL to the raw, playable video file.	meta
status	VideoStatus	Object describing the status attributes of a video.	
title	string	The video title or caption.	considered
updated_time	datetime	The last time the video or its caption was updated.	

Tabelle B.18: Feature-Kandidaten der *Edges* eines *Video Node* (nach [Fac17e])

Edge	Type	Description	Note
captions	list<VideoCaption>	Captions for the video.	considered
comments	list<Comment>	Users comments for the video.	partially considered
likes	list<Profile>	People who like this	
reactions	list<Profile>	People who reacted on this	
sharedposts	list<Post>	Posts that share this video.	
tags	list<Profile>	Users tagged in the video.	



## B.10 VideoList Node

Tabelle B.19: Feature-Kandidaten der *Fields* eines *VideoList Node* (nach [Fac17e])

Field	Type	Description	Note
creation_time	datetime	The time when the playlist was created	
description	string	Description of the playlist	considered
id	numeric string	Video playlist ID	meta
last_modified	datetime	The time when the contents of the playlist was last changed	
owner	User Page	Owner of the playlist	meta
thumbnail	string	Thumbnail of the playlist	
title	string	Title of the playlist	considered
videos_count	int32	Number of Videos in the playlist	meta

Tabelle B.20: Feature-Kandidaten der *Edges* eines *VideoList Node* (nach [Fac17e])

Edge	Type	Description	Note
videos	list<Video>	Videos in the playlist	considered



## Anhang C: Facebook-Profile relevanter Gruppierungen in Sachsen

Tabelle C.1: Facebook-Profile linksextremistischer Gruppierungen in Sachsen (Stand 10. September 2017, vgl. [Lan16, S. 34ff.])

Profile	Type	Facebook ID
New Kids Antifa Leipzig	Autonomous	166119460540635
Revolution Leipzig	Autonomous	311214862399136
The Future is unwritten Leipzig	Autonomous	196574267032483
Prisma Leipzig	Autonomous	440523829365325
Projekt Schuldenberg (Plauen)	Autonomous	1543684032598983
Anarchosyndikalistische Jugend Leipzig	Anarchical	220738771289582
Freie Arbeiterinnen- und Arbeiter-Union Dresden	Anarchical	1596300860593520
Freie Arbeiter-Union Leipzig	Anarchical	237016283134044
Freie Arbeiter-Union Chemnitz	Anarchical	311288999079471
Kommunistische Plattform der Partei „Die Linke“	Party	198998417111419
Sozialistische Deutsche Arbeiterjugend Leipzig	Party	692689324098142
Sozialistische Deutsche Arbeiterjugend Dresden	Party	478583515653388
Autonomal Versand	Music	1513479108926144
Comunidad de Libertad	Music	500065713361720
Sozialer Fehltritt (Chemnitz)	Music	299742233476532
Phase 2 (Leipzig)	Journal	185817254820500

Tabelle C.2: Facebook-Profile rechtsextremistischer Gruppierungen in Sachsen (Stand 10. September 2017, vgl. [Lan16, S. 5ff.])

Profile	Type	Facebook ID
NPD-Landesverband Sachsen	Party	495863713845799
NPD-Kreisverband Dresden	Party	283849118325985
NPD-Kreisverband Chemnitz	Party	238762262968312
NPD-Kreisverband Leipzig	Party	174812696054356
NPD-Kreisverband Zwickau-Westsachsen	Party	300610560106907
NPD-Kreisverband Niederschlesien-Oberlausitz	Party	820523998004929
NPD-Kreisverband Bautzen	Party	503020599847268
NPD-Kreisverband Meißen	Party	641695252616486
NPD-Kreisverband Vogtland	Party	251225418370226
NPD-Kreisverband Nordsachsen	Party	302566963092373
NPD-Kreisverband Sächsische Schweiz-Osterzgebirge	Party	207914785903716

Profile	Type	Facebook ID
NPD-Kreisverband Erzgebirge	Party	147875611942336
NPD-Kreisverband Mittelsachsen	Party	1446520022298636
NPD Altkreis Döbeln	Party	1502381580059089
NPD Leipzig Ortsgruppe Lossatal	Party	930616463628296
NPD-Ortsverband Sebnitz/Neustadt	Party	615718005158749
Haus Wieland (Riesa)	Party	1696224030626998
Haus Montag (Pirna)	Party	493185250774046
JN-Landesverband Sachsen	Party	261146433946102
JN Chemnitz	Party	613660632062821
JN Dresden	Party	1545585429027179
JN Mittelsachsen	Party	190208927807580
Ring Nationaler Frauen Sachsen	Party	313213892032516
Freundeskreis Udo Voigt	Party	377164779068207
Wir für Leipzig	Subculture	511080459004629
Freie Aktivisten Dresden	Subculture	1009299402472077
Kryptonit	Subculture	552613388181293
Blutzeugen	Music	1068096729934988
Brainwash	Music	171017036667305
Endless Struggle	Music	718856451474540
Hope for the Weak	Music	283445515115361
Killuminati	Music	231201543929744
Manson	Music	1541495942772336
Moshpit	Music	184297525038002
Neubeginn	Music	394417290620455
Paranoid	Music	610239862325051
Pionier	Music	707812745984967
Pro Patria	Music	161433637399165
Selbststeller	Music	269363823503920
Stereotyp	Music	578499135527142
Thematik 25	Music	429230053916868
Trendkiller	Music	873202402756898
Verboten	Music	1768062730147333
Volksnah	Music	1447356012227586
WUT	Music	342790055884155
Liedermacher Freilich Frei	Music	539884182811622
Liedermacher Freilich Frei	Music	1646427302295630
Verlag Nation und Wissen	Commerce	598368140178158
Dryve by Suizhyde Clothing	Commerce	133241330186715
Nordic Flame	Commerce	257913237633006
One People one Struggle Records	Commerce	1295335980494320
NMV-Versand	Commerce	430244987046518
Vergissmeinnicht 5.März Chemnitz	Event	911744845608548
1.Mai in Wurzen	Event	112015859192098
New Society 2004 (Chemnitz)	Other	667658503270879
Kaotic Chemnitz	Other	178791888908667
Europäische Aktion Sachsen	Other	546375068870543

## Anhang D: XML-Schema des Goldstandards

Listing D.1: XML-Schema des Goldstandards

```

<?xml version="1.0" encoding="UTF-8"?>
<xs:schema targetNamespace="http://www.bioforscher.de/stf/sna/goldstandard/"
  xmlns="http://www.bioforscher.de/stf/sna/goldstandard/"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  elementFormDefault="qualified">

  <xs:element name="page" type="pageType"></xs:element>

  <xs:complexType name="albumsType">
    <xs:sequence>
      <xs:element name="album" type="albumType" minOccurs="0" maxOccurs="unbounded"></xs:element>
    </xs:sequence>
  </xs:complexType>

  <xs:complexType name="albumType">
    <xs:sequence>
      <xs:element name="description" type="xs:string"></xs:element>
      <xs:element name="from" type="userType"></xs:element>
      <xs:element name="name" type="xs:string"></xs:element>
      <xs:element name="comments" type="commentsType"></xs:element>
      <xs:element name="photos" type="photosType"></xs:element>
      <xs:element name="picture" type="pictureType"></xs:element>
    </xs:sequence>
    <xs:attribute name="id" type="xs:string" use="required"></xs:attribute>
    <xs:attribute name="aprioriClassLabel" type="xs:string"></xs:attribute>
    <xs:attribute name="aprioriClassScore" type="xs:float"></xs:attribute>
    <xs:attribute name="calcClassLabel" type="xs:string"></xs:attribute>
    <xs:attribute name="calcClassScore" type="xs:float"></xs:attribute>
  </xs:complexType>

  <xs:complexType name="attachmentsType">
    <xs:sequence>
      <xs:element name="attachment" type="attachmentType" minOccurs="0" maxOccurs="unbounded"></xs:element>
    </xs:sequence>
  </xs:complexType>

  <xs:complexType name="attachmentType">
    <xs:attribute name="id" type="xs:string" use="required"></xs:attribute>
    <xs:attribute name="url" type="xs:string"></xs:attribute>
    <xs:attribute name="aprioriClassLabel" type="xs:string"></xs:attribute>
    <xs:attribute name="aprioriClassScore" type="xs:float"></xs:attribute>
    <xs:attribute name="calcClassLabel" type="xs:string"></xs:attribute>
    <xs:attribute name="calcClassScore" type="xs:float"></xs:attribute>
  </xs:complexType>

  <xs:complexType name="commentsType">
    <xs:sequence>
      <xs:element name="comment" type="commentType" minOccurs="0" maxOccurs="unbounded"></xs:element>
    </xs:sequence>
  </xs:complexType>

  <xs:complexType name="commentType">
    <xs:sequence>
      <xs:element name="from" type="userType"></xs:element>
      <xs:element name="message" type="xs:string"></xs:element>
    </xs:sequence>
    <xs:attribute name="id" type="xs:string" use="required"></xs:attribute>
    <xs:attribute name="commentCount" type="xs:int"></xs:attribute>
    <xs:attribute name="likeCount" type="xs:int"></xs:attribute>
  </xs:complexType>

```

```

<xs:attribute name="aprioriClassLabel" type="xs:string"></xs:attribute>
<xs:attribute name="aprioriClassScore" type="xs:float"></xs:attribute>
<xs:attribute name="calcClassLabel" type="xs:string"></xs:attribute>
<xs:attribute name="calcClassScore" type="xs:float"></xs:attribute>
</xs:complexType>

<xs:complexType name="eventsType">
  <xs:sequence>
    <xs:element name="event" type="eventType" minOccurs="0" maxOccurs="unbounded"></xs:element>
  </xs:sequence>
</xs:complexType>

<xs:complexType name="eventType">
  <xs:sequence>
    <xs:element name="category" type="xs:string"></xs:element>
    <xs:element name="description" type="xs:string"></xs:element>
    <xs:element name="name" type="xs:string"></xs:element>
    <xs:element name="owner" type="profileType"></xs:element>
    <xs:element name="comments" type="commentsType"></xs:element>
    <xs:element name="feed" type="postsType"></xs:element>
    <xs:element name="photos" type="photosType"></xs:element>
    <xs:element name="picture" type="pictureType"></xs:element>
  </xs:sequence>
  <xs:attribute name="id" type="xs:string" use="required"></xs:attribute>
  <xs:attribute name="attendingCount" type="xs:int"></xs:attribute>
  <xs:attribute name="declinedCount" type="xs:int"></xs:attribute>
  <xs:attribute name="interestedCount" type="xs:int"></xs:attribute>
  <xs:attribute name="maybeCount" type="xs:int"></xs:attribute>
  <xs:attribute name="noreplyCount" type="xs:int"></xs:attribute>
  <xs:attribute name="aprioriClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="aprioriClassScore" type="xs:float"></xs:attribute>
  <xs:attribute name="calcClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="calcClassScore" type="xs:float"></xs:attribute>
</xs:complexType>

<xs:complexType name="imagesType">
  <xs:sequence>
    <xs:element name="image" type="imageType" minOccurs="0" maxOccurs="unbounded"></xs:element>
  </xs:sequence>
</xs:complexType>

<xs:complexType name="imageType">
  <xs:attribute name="height" type="xs:int"></xs:attribute>
  <xs:attribute name="url" type="xs:string"></xs:attribute>
  <xs:attribute name="width" type="xs:int"></xs:attribute>
  <xs:attribute name="aprioriClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="aprioriClassScore" type="xs:float"></xs:attribute>
  <xs:attribute name="calcClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="calcClassScore" type="xs:float"></xs:attribute>
</xs:complexType>

<xs:complexType name="pageType">
  <xs:sequence>
    <xs:element name="about" type="xs:string"></xs:element>
    <xs:element name="bestPage" type="pageType"></xs:element>
    <xs:element name="category" type="xs:string"></xs:element>
    <xs:element name="categoryList" type="xs:string"></xs:element>
    <xs:element name="description" type="xs:string"></xs:element>
    <xs:element name="featuredVideo" type="videoType"></xs:element>
    <xs:element name="generalInfo" type="xs:string"></xs:element>
    <xs:element name="impressum" type="xs:string"></xs:element>
    <xs:element name="name" type="xs:string"></xs:element>
    <xs:element name="userName" type="xs:string"></xs:element>
    <xs:element name="albums" type="albumsType"></xs:element>
    <xs:element name="events" type="eventsType"></xs:element>
  </xs:sequence>

```

```

    <xs:element name="photos" type="photosType"></xs:element>
    <xs:element name="picture" type="pictureType"></xs:element>
    <xs:element name="posts" type="postsType"></xs:element>
    <xs:element name="videoLists" type="videoListsType"></xs:element>
    <xs:element name="videos" type="videosType"></xs:element>
  </xs:sequence>
  <xs:attribute name="id" type="xs:string" use="required"></xs:attribute>
  <xs:attribute name="fanCount" type="xs:int"></xs:attribute>
  <xs:attribute name="isCommunityPage" type="xs:boolean"></xs:attribute>
  <xs:attribute name="link" type="xs:string"></xs:attribute>
  <xs:attribute name="overallStarRating" type="xs:float"></xs:attribute>
  <xs:attribute name="ratingCount" type="xs:int"></xs:attribute>
  <xs:attribute name="talkingAboutCount" type="xs:int"></xs:attribute>
  <xs:attribute name="verificationStatus" type="xs:string"></xs:attribute>
  <xs:attribute name="website" type="xs:string"></xs:attribute>
  <xs:attribute name="aprioriClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="aprioriClassScore" type="xs:float"></xs:attribute>
  <xs:attribute name="calcClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="calcClassScore" type="xs:float"></xs:attribute>
</xs:complexType>

<xs:complexType name="photosType">
  <xs:sequence>
    <xs:element name="photo" type="photoType" minOccurs="0" maxOccurs="unbounded"></xs:element>
  </xs:sequence>
</xs:complexType>

<xs:complexType name="photoType">
  <xs:sequence>
    <xs:element name="album" type="albumType"></xs:element>
    <xs:element name="from" type="profileType"></xs:element>
    <xs:element name="images" type="imagesType"></xs:element>
    <xs:element name="name" type="xs:string"></xs:element>
    <xs:element name="comments" type="commentsType"></xs:element>
  </xs:sequence>
  <xs:attribute name="id" type="xs:string" use="required"></xs:attribute>
  <xs:attribute name="height" type="xs:int"></xs:attribute>
  <xs:attribute name="link" type="xs:string"></xs:attribute>
  <xs:attribute name="width" type="xs:int"></xs:attribute>
  <xs:attribute name="aprioriClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="aprioriClassScore" type="xs:float"></xs:attribute>
  <xs:attribute name="calcClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="calcClassScore" type="xs:float"></xs:attribute>
</xs:complexType>

<xs:complexType name="pictureType">
  <xs:attribute name="height" type="xs:int"></xs:attribute>
  <xs:attribute name="url" type="xs:string"></xs:attribute>
  <xs:attribute name="width" type="xs:int"></xs:attribute>
  <xs:attribute name="aprioriClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="aprioriClassScore" type="xs:float"></xs:attribute>
  <xs:attribute name="calcClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="calcClassScore" type="xs:float"></xs:attribute>
</xs:complexType>

<xs:complexType name="postsType">
  <xs:sequence>
    <xs:element name="post" type="postType" maxOccurs="unbounded" minOccurs="0"></xs:element>
  </xs:sequence>
</xs:complexType>

<xs:complexType name="postType">
  <xs:sequence>
    <xs:element name="description" type="xs:string"></xs:element>
    <xs:element name="from" type="profileType"></xs:element>
  </xs:sequence>

```

```

    <xs:element name="message" type="xs:string"></xs:element>
    <xs:element name="name" type="xs:string"></xs:element>
    <xs:element name="attachments" type="attachmentsType"></xs:element>
    <xs:element name="comments" type="commentsType"></xs:element>
  </xs:sequence>
  <xs:attribute name="id" type="xs:string" use="required"></xs:attribute>
  <xs:attribute name="link" type="xs:string"></xs:attribute>
  <xs:attribute name="permalinkUrl" type="xs:string"></xs:attribute>
  <xs:attribute name="shareCount" type="xs:int"></xs:attribute>
  <xs:attribute name="statusType" type="statusTypeType"></xs:attribute>
  <xs:attribute name="type" type="typeType"></xs:attribute>
  <xs:attribute name="aprioriClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="aprioriClassScore" type="xs:float"></xs:attribute>
  <xs:attribute name="calcClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="calcClassScore" type="xs:float"></xs:attribute>
</xs:complexType>

<xs:complexType name="profileType">
  <xs:choice>
    <xs:element name="page" type="pageType"></xs:element>
    <xs:element name="user" type="userType"></xs:element>
  </xs:choice>
</xs:complexType>

<xs:simpleType name="statusTypeType">
  <xs:restriction base="xs:string">
    <xs:enumeration value="added_photos" />
    <xs:enumeration value="added_video" />
    <xs:enumeration value="app_created_story" />
    <xs:enumeration value="approved_friend" />
    <xs:enumeration value="created_event" />
    <xs:enumeration value="created_group" />
    <xs:enumeration value="created_note" />
    <xs:enumeration value="mobile_status_update" />
    <xs:enumeration value="published_story" />
    <xs:enumeration value="shared_story" />
    <xs:enumeration value="tagged_in_photo" />
    <xs:enumeration value="wall_post" />
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="typeType">
  <xs:restriction base="xs:string">
    <xs:enumeration value="datetime" />
    <xs:enumeration value="link" />
    <xs:enumeration value="offer" />
    <xs:enumeration value="photo" />
    <xs:enumeration value="status" />
    <xs:enumeration value="video" />
  </xs:restriction>
</xs:simpleType>

<xs:complexType name="userType">
  <xs:attribute name="id" type="xs:string" use="required"></xs:attribute>
  <xs:attribute name="firstName" type="xs:string"></xs:attribute>
  <xs:attribute name="lastName" type="xs:string"></xs:attribute>
  <xs:attribute name="middleName" type="xs:string"></xs:attribute>
  <xs:attribute name="name" type="xs:string"></xs:attribute>
  <xs:attribute name="aprioriClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="aprioriClassScore" type="xs:float"></xs:attribute>
  <xs:attribute name="calcClassLabel" type="xs:string"></xs:attribute>
  <xs:attribute name="calcClassScore" type="xs:float"></xs:attribute>
</xs:complexType>

<xs:complexType name="videoListsType">

```



```
<xs:sequence>
  <xs:element name="videoList" type="videoListType" minOccurs="0" maxOccurs="unbounded"/></xs:element>
</xs:sequence>
</xs:complexType>

<xs:complexType name="videoListType">
  <xs:sequence>
    <xs:element name="description" type="xs:string"/></xs:element>
    <xs:element name="owner" type="profileType"/></xs:element>
    <xs:element name="title" type="xs:string"/></xs:element>
    <xs:element name="videos" type="videosType"/></xs:element>
  </xs:sequence>
  <xs:attribute name="id" type="xs:string" use="required"/></xs:attribute>
  <xs:attribute name="videosCount" type="xs:int"/></xs:attribute>
  <xs:attribute name="aprioriClassLabel" type="xs:string"/></xs:attribute>
  <xs:attribute name="aprioriClassScore" type="xs:float"/></xs:attribute>
  <xs:attribute name="calcClassLabel" type="xs:string"/></xs:attribute>
  <xs:attribute name="calcClassScore" type="xs:float"/></xs:attribute>
</xs:complexType>

<xs:complexType name="videosType">
  <xs:sequence>
    <xs:element name="video" type="videoType" minOccurs="0" maxOccurs="unbounded"/></xs:element>
  </xs:sequence>
</xs:complexType>

<xs:complexType name="videoType">
  <xs:sequence>
    <xs:element name="contentCategory" type="xs:string"/></xs:element>
    <xs:element name="contentTags" type="xs:string"/></xs:element>
    <xs:element name="description" type="xs:string"/></xs:element>
    <xs:element name="from" type="profileType"/></xs:element>
    <xs:element name="title" type="xs:string"/></xs:element>
    <xs:element name="captions" type="xs:string"/></xs:element>
    <xs:element name="comments" type="commentsType"/></xs:element>
  </xs:sequence>
  <xs:attribute name="id" type="xs:string" use="required"/></xs:attribute>
  <xs:attribute name="format" type="xs:string"/></xs:attribute>
  <xs:attribute name="length" type="xs:float"/></xs:attribute>
  <xs:attribute name="permalinkUrl" type="xs:string"/></xs:attribute>
  <xs:attribute name="source" type="xs:string"/></xs:attribute>
  <xs:attribute name="aprioriClassLabel" type="xs:string"/></xs:attribute>
  <xs:attribute name="aprioriClassScore" type="xs:float"/></xs:attribute>
  <xs:attribute name="calcClassLabel" type="xs:string"/></xs:attribute>
  <xs:attribute name="calcClassScore" type="xs:float"/></xs:attribute>
</xs:complexType>

</xs:schema>
```



## Literaturverzeichnis

- [AC12] ATSERIAS, Jordi; CODINA, Joan: What Is the Text of a Tweet? In: *Proceedings of the @ NLP can u tag #usergeneratedcontent* (2012), S. 29–33
- [Agg14] AGGARWAL, Charu C.: *Data Classification: Algorithms and Applications*. CRC Press, 2014
- [Ale17a] ALEXA INTERNET, INC.: *The Top 500 Sites on the Web*. <http://www.alex.com/topsites> [zuletzt abgerufen am 16. August 2017]. August 2017
- [Ale17b] ALEXA INTERNET, INC.: *Top Sites in Germany*. <http://www.alex.com/topsites/countries/DE> [zuletzt abgerufen am 16. August 2017]. August 2017
- [Atz12] ATZMUELLER, Martin: Mining Social Media: Key Players, Sentiments, and Communities. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2 (2012), September, Nr. 5, S. 411–419
- [AZ12] AGGARWAL, Charu C. (Hrsg.); ZHAI, ChengXiang (Hrsg.): *Mining Text Data*. Boston, MA : Springer US, 2012
- [BE07] BOYD, Danah M.; ELLISON, Nicole B.: Social Network Sites: Definition, History, and Scholarship. In: *Journal of Computer-Mediated Communication* 13 (2007), Oktober, Nr. 1, S. 210–230
- [BH83] BARNES, John A.; HARARY, Frank: Graph Theory in Network Analysis. In: *Social Networks* 5 (1983), Nr. 2, S. 235–244
- [BL97] BLUM, Avrim L.; LANGLEY, Pat: Selection of Relevant Features and Examples in Machine Learning. In: *Artificial Intelligence* 97 (1997), Dezember, Nr. 1, S. 245–271
- [BM11] BYRNE, James; MARX, Gary: Technological Innovations in Crime Prevention and Policing. A Review of the Research on Implementation and Impact. In: *Journal of Police Studies* 20 (2011), Nr. 3, S. 17–40
- [BMBL09] BORGATTI, S. P.; MEHRA, A.; BRASS, D. J.; LABIANCA, G.: Network Analysis in the Social Sciences. In: *Science* 323 (2009), Februar, Nr. 5916, S. 892–895

- [BP98] BRIN, Sergey; PAGE, Lawrence: The Anatomy of a Large-Scale Hypertextual Web Search Engine. In: *Computer Networks and ISDN Systems* 30 (1998), S. 107–117
- [BR96] BORDO, Michael D.; ROCKOFF, Hugh: The Gold Standard as a “Good Housekeeping Seal of Approval”. In: *The Journal of Economic History* 56 (1996), Nr. 2, S. 389–428
- [BR14] BONTCHEVA, Kalina; ROUT, Dominic: Making Sense of Social Media Streams through Semantics: A Survey. In: *Semantic Web* 5 (2014), Nr. 5, S. 373–403
- [Bun16] BUNDESMINISTERIUM DER JUSTIZ UND FÜR VERBRAUCHERSCHUTZ: *Pressemitteilungen: Löschung von strafbaren Hasskommentaren im Netz noch nicht ausreichend*. September 2016
- [Bun17] BUNDESKRIMINALAMT: *Presseinformation: Neues Instrument zur Risikobewertung von potentiellen Gewaltstraftätern*. Februar 2017
- [Bö11] BÖHM, María L.: *Der Gefährder und das Gefährdungsrecht: eine rechtssoziologische Analyse am Beispiel der Urteile des Bundesverfassungsgerichts über die nachträgliche Sicherungsverwahrung und die akustische Wohnraumüberwachung*. Göttingen : Univ.-Verl. Göttingen, 2011 (Göttinger Studien zu den Kriminalwissenschaften 15)
- [CDRT07] CARLEY, Kathleen M.; DIESNER, Jana; REMINGA, Jeffrey; TSVETOVAT, Maksim: Toward an Interoperable Dynamic Network Analysis Toolkit. In: *Decision Support Systems* 43 (2007), August, Nr. 4, S. 1324–1347
- [Cen10] CENTOLA, D.: The Spread of Behavior in an Online Social Network Experiment. In: *Science* 329 (2010), September, Nr. 5996, S. 1194–1197
- [DACP15] DUNBAR, R.I.M.; ARNABOLDI, Valerio; CONTI, Marco; PASSARELLA, Andrea: The Structure of Online Social Networks Mirrors Those in the Offline World. In: *Social Networks* 43 (2015), Oktober, S. 39–47
- [Deu16] DEUTSCHE BUNDESREGIERUNG: *Einstufung einer Anti-Atom-Aktivistin als „Gefährder“ und „relevante Person“: Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Hubertus Zdebel, Ulla Jelpke, Caren Lay, weiterer Abgeordneter und der Fraktion DIE LINKE – Drucksache 18/10176*. Deutscher Bundestag. Drucksache 18/10340. November 2016
- [Deu17a] DEUTSCHE BUNDESREGIERUNG: *Gefährder in Deutschland: Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Ulla Jelpke,*

- Martina Renner, Dr. André Hahn, weiterer Abgeordneter und der Fraktion DIE LINKE – Drucksache 18/11064. Deutscher Bundestag. Drucksache 18/11369. März 2017*
- [Deu17b] DEUTSCHE BUNDESREGIERUNG: *Problematik des Gefährder-Begriffs (Nachfrage zur Antwort der Bundesregierung auf die Kleine Anfrage auf Bundestagsdrucksache 18/11369): Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Ulla Jelpke, Frank Tempel, Dr. André Hahn, weiterer Abgeordneter und der Fraktion DIE LINKE – Drucksache 18/11959. Deutscher Bundestag. Drucksache 18/12196. Februar 2017*
- [Deu17c] DEUTSCHE STIFTUNG WELTBEVÖLKERUNG: *Wie Viele Menschen Leben Auf Der Erde? <https://www.dsw.org> [zuletzt abgerufen am 16. August 2017]. August 2017*
- [DS10] DROZDOVA, Katya; SAMOILOV, Michael: Predictive Analysis of Concealed Social Network Activities Based on Communication Technology Choices: Early-Warning Detection of Attack Signals from Terrorist Organizations. In: *Computational and Mathematical Organization Theory* 16 (2010), März, Nr. 1, S. 61–88
- [EBGC09] ERÉTÉO, Guillaume; BUFFA, Michel; GANDON, Fabien; CORBY, Olivier: Analysis of a Real Online Social Network Using Semantic Web Frameworks. In: *The Semantic Web-ISWC 2009* (2009), S. 180–195
- [Ecl16] ECLIPSE FOUNDATION: *Eclipse XML Editors and Tools. <https://marketplace.eclipse.org/content/eclipse-xml-editors-and-tools-0> [zuletzt abgerufen am 5. September 2017]. Juni 2016*
- [Ecl17] ECLIPSE FOUNDATION: *Eclipse Platform and IDE. <http://www.eclipse.org> [zuletzt abgerufen am 4. Juli 2017]. Juli 2017*
- [Egb17] EGBERT, Simon: Siegeszug Der Algorithmen? – Predictive Policing im deutschsprachigen Raum. In: *Aus Politik und Zeitgeschichte: Innere Sicherheit. Bundeszentrale für politische Bildung* 67 (2017), August, Nr. 32-33, S. 17–23
- [EPK<sup>+</sup>11] EDOSOMWAN, Simeon; PRAKASAN, Sitalaskshmi K.; KOUAME, Doriane; WATSON, Jonelle; SEYMOUR, Tom: The History of Social Media and Its Impact on Business. In: *Journal of Applied Management and entrepreneurship* 16 (2011), Nr. 3, S. 79
- [Fac17a] FACEBOOK INC.: *Eine Community von 30 Millionen: Facebook sagt Danke. <https://de.newsroom.fb.com/news/2017/06/eine-community->*

- von-30-millionen-facebook-sagt-danke/ [zuletzt abgerufen am 16. August 2017]. Juni 2017*
- [Fac17b] FACEBOOK INC.: *Facebook Hilfebereich*. <https://www.facebook.com/help/> [zuletzt abgerufen am 15. August 2017]. August 2017
- [Fac17c] FACEBOOK INC.: *Facebook Reports Second Quarter 2017 Results*. <https://investor.fb.com/investor-news/press-release-details/2017/Facebook-Reports-Second-Quarter-2017-Results/default.aspx> [zuletzt abgerufen am 16. August 2017]. Juli 2017
- [Fac17d] FACEBOOK INC.: *SEC Filings Details 4Q 2016. FORM 10-Q*. <https://investor.fb.com/financials/sec-filings-details/default.aspx?FilingId=12041947> [zuletzt abgerufen am 17. August 2017]. Mai 2017
- [Fac17e] FACEBOOK INC.: *The Graph API Documentation: Graph API Reference*. <https://developers.facebook.com/docs/graph-api/reference> [zuletzt abgerufen am 4. Juli 2017]. Juli 2017
- [Fac17f] FACEBOOK INC.: *The Graph API Documentation: Rate Limiting on the Graph API*. <https://developers.facebook.com/docs/graph-api/advanced/rate-limiting> [zuletzt abgerufen am 4. Juli 2017]. Juli 2017
- [FD84] FARARO, Thomas J.; DOREIAN, Patrick: Tripartite Structural Analysis: Generalizing the Breiger-Wilson Formalism. In: *Social Networks* 6 (1984), Nr. 2, S. 141–175
- [For03] FORMAN, George: An Extensive Empirical Study of Feature Selection Metrics for Text Classification. In: *Journal of machine learning research* 3 (2003), Nr. Mar, S. 1289–1305
- [GL12] GUNDECHA, Pritam; LIU, Huan: Mining Social Media: A Brief Introduction. In: *2012 TutORials in Operations Research*. INFORMS, Oktober 2012, S. 1–17
- [Gua13] GUARDIAN NEWS AND MEDIA LIMITED: *Interview des Guardian mit Edward Snowden*. <https://www.theguardian.com/world/2013/jun/09/nsa-whistleblower-edward-snowden-why> [zuletzt abgerufen am 11. September 2017]. Juni 2013
- [HKP12] HEIDEMANN, Julia; KLIER, Mathias; PROBST, Florian: Online Social Networks: A Survey of a Global Phenomenon. In: *Computer Networks* 56

- (2012), Dezember, Nr. 18, S. 3866–3878
- [IKT05] IKONOMAKIS, M; KOTSIANTIS, S; TAMPAKAS, V: Text Classification Using Machine Learning Techniques. In: *WSEAS transactions on computers* 4 (2005), Nr. 8, S. 966–974
- [Juc12] JUCKETT, David: A Method for Determining the Number of Documents Needed for a Gold Standard Corpus. In: *Journal of Biomedical Informatics* 45 (2012), Juni, Nr. 3, S. 460–470
- [KHMS11] KIETZMANN, Jan H.; HERMKENS, Kristopher; MCCARTHY, Ian P.; SILVESTRE, Bruno S.: Social Media? Get Serious! Understanding the Functional Building Blocks of Social Media. In: *Business Horizons* 54 (2011), Mai, Nr. 3, S. 241–251
- [Kos14] KOSINSKI, Michal: Measurement and Prediction of Individual and Group Differences in the Digital Environment. In: *Department of Psychology University of Cambridge* (2014)
- [Kot07] KOTSIANTIS, SB: Supervised Machine Learning: A Review of Classification Techniques. In: *Informatica* 31 (2007), S. 249–268
- [KR00] KITTLER, Josef (Hrsg.); ROLI, Fabio (Hrsg.): *Multiple Classifier Systems: First International Workshop, MCS 2000, Cagliari, Italy, June 21-23, 2000: Proceedings*. Berlin ; New York : Springer, 2000 (Lecture notes in computer science 1857)
- [Kre17] KRETSCHMANN, Andrea: Soziale Tatsachen – Eine Wissenssoziologische Perspektive Auf Den „Gefährder“. In: *Aus Politik und Zeitgeschichte: Innere Sicherheit. Bundeszentrale für politische Bildung* 67 (2017), August, Nr. 32-33, S. 11–16
- [KS16] KAUR, Ravneet; SINGH, Sarbjeet: A Survey of Data Mining and Social Network Analysis Based Anomaly Detection Techniques. In: *Egyptian Informatics Journal* 17 (2016), Juli, Nr. 2, S. 199–216
- [Lan16] LANDESAMT FÜR VERFASSUNGSSCHUTZ SACHSEN: *Internetatlas 2016. Aktivitäten von Extremisten im World Wide Web*. Juli 2016
- [Lan17a] LANDESAMT FÜR VERFASSUNGSSCHUTZ SACHSEN: *Sächsischer Verfassungsschutzbericht 2016*. April 2017
- [Lan17b] LANDESREGIERUNG BRANDENBURG: *Einstufungskriterien Für „Gefährder“: Antwort Der Landesregierung Auf Die Kleine Anfrage Nr. 2496 Der*

*Abgeordneten Ursula Nonnemacher Und Benjamin Raschke Der Fraktion BÜNDNIS 90/DIE GRÜNEN – Drucksache 6/6089. Landtag Brandenburg. Drucksache 6/6272. März 2017*

- [LCW10] LEE, Kyumin; CAVERLEE, James; WEBB, Steve: Uncovering Social Spammers: Social Honeypots + Machine Learning, ACM Press, 2010, S. 435
- [LLS00] LIM, Tjen-Sien; LOH, Wei-Yin; SHIH, Yu-Shan: A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms. In: *Machine learning* 40 (2000), Nr. 3, S. 203–228
- [LWC14] LI, Rui; WANG, Chi; CHANG, Kevin Chen-Chuan: User Profiling in an Ego Network: Co-Profiling Attributes and Relationships, ACM Press, 2014, S. 819–830
- [Mar16] MARÉCHAL, Nathalie: Automation, Algorithms, and Politics| When Bots Tweet: Toward a Normative Framework for Bots on Social Networking Sites (Feature). In: *International Journal of Communication* 10 (2016), S. 10
- [MBR12] MAYNARD, Diana; BONTCHEVA, Kalina; ROUT, Dominic: Challenges in Developing Opinion Mining Tools for Social Media. In: *Proceedings of the@ NLP can u tag #usergeneratedcontent* (2012), S. 15–22
- [MBW13] MELAMED, David; BREIGER, Ronald L.; WEST, A J.: Community Structure in Multi-Mode Networks: Applying an Eigenspectrum Approach. In: *Official Journal of the International Network for Social Network Analysts* 33 (2013), Nr. 1, S. 18–23
- [Med17] MEDILEXICON INTERNATIONAL LTD: *MediLexicon. Gold Standard.* <http://www.medilexicon.com/dictionary/38019> [zuletzt abgerufen am 4. Juli 2017]. Juli 2017
- [MGN12] MUNOZ-GARCIA, Oscar; NAVARRO, Carlos: Comparing User-Generated Content Published in Different Social Media Sources. In: *Workshop NLP Can u Tag #usergenerated-Content*, 2012, S. 1–8
- [Mor34] MORENO, Jacob L.: Who Shall Survive?: A New Approach to the Problem of Human Interrelations. (1934)
- [MW11] MARIN, Alexandra; WELLMAN, Barry: Social Network Analysis: An Introduction. In: *The SAGE handbook of social network analysis* 11 (2011)
- [MyP15] MYPERSONALITY PROJECT: *Download Databases.*



- [http://mypersonality.org/wiki/doku.php?id=download\\_databases](http://mypersonality.org/wiki/doku.php?id=download_databases) [zuletzt abgerufen am 4. September 2017]. 2015
- [PB16] PARVATHY, G.; BINDHU, J. S.: A Probabilistic Generative Model for Mining Cybercriminal Network from Online Social Media: A Review. In: *International Journal of Computer Applications* 134 (2016), Nr. 14
- [PFML16] POZZI, Federico A.; FERSINI, Elisabetta; MESSINA, Enza; LIU, Bing: *Sentiment Analysis in Social Networks*. 1. Morgan Kaufmann, 2016
- [PKVS12] PAPADOPOULOS, Symeon; KOMPATSIARIS, Yiannis; VAKALI, Athena; SPYRIDONOS, Ploutarchos: Community Detection in Social Media: Performance and Application Considerations. In: *Data Mining and Knowledge Discovery* 24 (2012), Mai, Nr. 3, S. 515–554
- [PP11] PENNACCHIOTTI, Marco; POPESCU, Ana-Maria: A Machine Learning Approach to Twitter User Classification. In: *lcwsm* 11 (2011), Nr. 1, S. 281–288
- [PS13] PUSTEJOVSKY, J.; STUBBS, Amber: *Natural Language Annotation for Machine Learning*. Sebastopol, CA : O'Reilly Media, 2013
- [Raf17] RAFFELSBERGER, Peter: *Online XML Schema Viewer*. <http://www.xml-tools.net> [zuletzt abgerufen am 5. September 2017]. Juli 2017
- [Rei16] REITMAIER, Tobias: Aktives Lernen für Klassifikationsprobleme unter der Nutzung von Strukturinformationen. In: *Steffen Hölldobler et. Al. (Hrsg.): Ausgezeichnete Informatikdissertationen 2015*. Bonn : Gesellschaft für Informatik, 2016 (Lecture Notes in Informatics), S. 239–248
- [RGG16] REUTER, Christian; GEILEN, Gordian; GELLERT, Robin: Sicherheit vs. Privatsphäre: Zur Akzeptanz von Überwachung in Sozialen Medien Im Kontext von Terrorkrisen. In: *GI-Jahrestagung*, 2016, S. 1759–1773
- [RN16] RAMANI, R. G.; NANCY, P.: Discovery of Classification Rules in Prediction of Applications Usage in Social Network Data (Facebook Application Data) Using Data Mining Algorithms. In: *International Journal on Human Machine Interaction* 1 (2016), Nr. 1, S. 28–36
- [Ro09] ROWE, Matthew; OTHERS: Interlinking Distributed Social Graphs. In: *LDOW*, 2009
- [ROR16] ROBERTZ, Prof. Dr. Frank J.; OKSANEN, Prof. Dr. A.; RÄSÄNEN, Prof. Dr. P.: Onlinebasierter Hass. In: *Viktimisierung junger Menschen im Internet*.

Springer Fachmedien Wiesbaden, 2016, S. 7–12

- [Sac13] *SächsPolG (idF vom 17.12.2013) § 1*. Dezember 2013
- [SBHL17] SPRANGER, Michael; BECKER, Sven; HEINKE, Florian; LABUDDE, Dirk: The Infiltration Game: Artificial Immune System for the Exploitation of Crime Relevant Information in Social Networks. In: *International Conference on Advances in Information Mining and Management (IMMM)*. Volume: 7. Venice, Juni 2017
- [Sco11] SCOTT, John: Social Network Analysis: Developments, Advances, and Prospects. In: *Social Network Analysis and Mining* 1 (2011), Januar, Nr. 1, S. 21–26
- [Seb01] SEBASTIANI, Fabrizio: Machine Learning in Automated Text Categorization. (2001), Oktober
- [SHAL16] SPRANGER, Michael; HEINKE, Florian; APPELT, Luisa; LABUDDE, Dirk: MoNA: Automated Identification of Evidence in Forensic Short Messages. In: *International Journal On Advances in Security* (2016)
- [SHGL15] SPRANGER, Michael; HEINKE, Florian; GRUNERT, Steffen; LABUDDE, Dirk: Towards Predictive Policing: Knowledge-Based Monitoring of Social Networks. Brussels, Juni 2015
- [SL13] SPRANGER, Michael; LABUDDE, Dirk: Semantic Tools for Forensics: Approaches in Forensic Text Analysis. In: *Proc. 3rd. International Conference on Advances in Information Management and Mining (IMMM), IARIA. ThinkMind Library*, 2013, S. 97–100
- [SL14a] SPRANGER, Michael; LABUDDE, D: Semantic Tools for Forensics: Towards Finding Evidence in Short Messages. In: *The Fourth International Conference on Advances in Information Mining and Management*, 2014, S. 1–4
- [SL14b] SPRANGER, Michael; LABUDDE, Dirk: Towards Establishing an Expert System for Forensic Text Analysis. In: *International Journal on Advances in Intelligent Systems* 7 (2014), Nr. 1/2, S. 247–256
- [SL17] SPRANGER, Michael; LABUDDE, Dirk: Textforensik. In: *Forensik in der digitalen Welt*. Springer Spektrum, Berlin, Heidelberg, 2017, S. 167–198
- [SRML09] SUN, Eric; ROSENN, Itamar; MARLOW, Cameron; LENTO, Thomas M.: Gesundheit! Modeling Contagion through Facebook News Feed. In: *ICWSM*, 2009

- [SSH<sup>+</sup>12] SPRANGER, Michael; SCHILDBACH, Stefan; HEINKE, Florian; GRUNERT, Steffen; LABUDDE, Dirk: *Semantic Tools for Forensics: A Highly Adaptable Framework*. 2012
- [Sta17] STATISTA GMBH: *Most Famous Social Network Sites Worldwide as of August 2017, Ranked by Number of Active Users (in Millions)*. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> [zuletzt abgerufen am 16. August 2017]. Januar 2017
- [StP17a] *StPO (idF Vom 17.08.2017) § 100a*. August 2017
- [StP17b] *StPO (idF Vom 17.08.2017) § 152 II*. August 2017
- [SYBT15] SILVESTRI, Giuseppe; YANG, Jie; BOZZON, Alessandro; TAGARELLI, Andrea: Linking Accounts across Social Networks: The Case of StackOverflow, Github and Twitter. In: *KDWeb*, 2015, S. 41–52
- [TAH<sup>+</sup>16] THURASINGHAM, Bhavani; ABROL, Satyen; HEATHERLY, Raymond; KANTARCIOGLU, Murat; KHADILKAR, Vaibhav; KHAN, Latifur: *Analyzing and Securing Social Networks*. Boston, MA, USA : Auerbach Publications, 2016
- [TG16] TAYEBI, Mohammad A.; GLÄSSER, Uwe: Social Network Analysis in Predictive Policing. In: *Social Network Analysis in Predictive Policing*. Springer, Cham, 2016 (Lecture Notes in Social Networks), S. 7–14
- [THS11] TRAPPMANN, Mark; HUMMELL, Hans J.; SODEUR, Wolfgang: *Strukturanalyse sozialer Netzwerke: Konzepte, Modelle, Methoden*. Wiesbaden : VS-Verl., 2011
- [TL10] TANG, Lei; LIU, Huan: Community Detection and Mining in Social Media. In: *Synthesis Lectures on Data Mining and Knowledge Discovery 2* (2010), Januar, Nr. 1, S. 1–137
- [WAMDP14] WISSLER, Lars; ALMASHRAEE, Mohammed; MONETT DÍAZ, Dagmar; PASCHKE, Adrian: The Gold Standard in Corpus Annotation. In: *IEEE GSC*, 2014
- [WCDO09] WEINSTEIN, Clifford; CAMPBELL, William; DELANEY, Brian; O’LEARY, Gerald: Modeling and Detection Techniques for Counter-Terror Social Network Analysis and Intent Recognition, *IEEE*, März 2009, S. 1–16
- [Wis08] WISSENSCHAFTLICHE DIENSTE DES DEUTSCHEN BUNDESTAGES: *Aktuel-*

ler Begriff: „Gefährder“. Nr. 36/08. Juli 2008

- [Wis17] WISSENSCHAFTLICHE DIENSTE DES DEUTSCHEN BUNDESTAGES: *Legal-definition des Begriffes „Gefährder“*. Aktenzeichen: WD 3 - 3000 - 046/17. Februar 2017
- [WKA16] WANI, Suheel Y.; KIRMANI, Mudasir M.; ANSARULLA, Syed I.: Prediction of Fake Profiles on Facebook Using Supervised Machine Learning Techniques-A Theoretical Model. In: *International Journal of Computer Science and Information Technologies* 7 (2016), Nr. 4, S. 1735–1738
- [Wor12a] WORLD WIDE WEB CONSORTIUM (W3C): *W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures*. <https://www.w3.org/TR/xmlschema11-1/> [zuletzt abgerufen am 10. September 2017]. April 2012
- [Wor12b] WORLD WIDE WEB CONSORTIUM (W3C): *W3C XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes*. <https://www.w3.org/TR/xmlschema11-2/> [zuletzt abgerufen am 10. September 2017]. April 2012
- [Wor16a] WORLD WIDE WEB CONSORTIUM (W3C): *Extensible Markup Language (XML)*. <https://www.w3.org/XML/> [zuletzt abgerufen am 9. September 2017]. Oktober 2016
- [Wor16b] WORLD WIDE WEB CONSORTIUM (W3C): *XML Schema*. <https://www.w3.org/XML/Schema> [zuletzt abgerufen am 10. September 2017]. April 2016
- [XZHC13] XIA, Rui; ZONG, Chengqing; HU, Xuelei; CAMBRIA, Erik: Feature Ensemble plus Sample Selection: Domain Adaptation for Sentiment Classification. In: *IEEE Intelligent Systems* 28 (2013), Nr. 3, S. 10–18
- [YK12] YU, Sheng; KAK, Subhash: A Survey of Prediction Using Social Media. (2012), März
- [ZL13] ZAFARANI, Reza; LIU, Huan: Connecting Users across Social Media Sites: A Behavioral-Modeling Approach. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2013, S. 41–49
- [ZM12] ZHANG, Cha (Hrsg.); MA, Yunqian (Hrsg.): *Ensemble Machine Learning*. Boston, MA : Springer US, 2012

- [ZM16] ZHAI, ChengXiang; MASSUNG, Sean: *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. First edition. New York : Association for Computing Machinery, 2016 (ACM books 12)



## Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Mittweida, 13. September 2017

Stefan Pabst